Cochrane Library

Cochrane Database of Systematic Reviews

Exercise programmes for ankylosing spondylitis (Review)

Regnaux JP, Davergne T, Palazzo C, Roren A, Rannou F, Boutron I, Lefevre-Colau MM

Regnaux JP, Davergne T, Palazzo C, Roren A, Rannou F, Boutron I, Lefevre-Colau MM.
Exercise programmes for ankylosing spondylitis.
Cochrane Database of Systematic Reviews 2019, Issue 10. Art. No.: CD011321.
DOI: 10.1002/14651858.CD011321.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Exercise programmes for ankylosing spondylitis

Jean-Philippe Regnaux[1,2,3], Thomas Davergne[4], Clémence Palazzo[5], Alexandra Roren[6], François Rannou[7], Isabelle Boutron[4,7,8,9], Marie-Martine Lefevre-Colau[6]

[1]Université Rennes, EHESP, F-3500 Rennes, France. [2] Université de Paris , Equipe Epi Ageing, Inserm, CRESS- UMR1153, Paris, France. [2]EHESP School of Public Health, Rennes, France. [3]Centre Cochrane Français, Paris, France. [4]French Cochrane Center, Paris, France. [5]Department of Physical Medicine and Rehabilitation, Assistance Publique des Hôpitaux de Paris (AP-HP.5), Hôpital Cochin, Paris, France. [6]ECaMO team, INSERM U1153, Paris, France. [7]Paris Descartes University, Sorbonne Paris Cité, Faculté de Médecine, Paris, France. [8]METHODS team, INSERM U1153, Paris, France. [9]Centre d'Épidémiologie Clinique, AP-HP (Assistance Publique des Hôpitaux de Paris), Hôpital Hôtel Dieu, Paris, France

Contact address: Jean-Philippe Regnaux, Université Rennes, EHESP, F-3500 Rennes, France. [2] Université de Paris , Equipe Epi Ageing, Inserm, CRESS- UMR1153, Paris, France. jean-philippe.regnaux@ehesp.fr.

Editorial group: Cochrane Musculoskeletal Group.
Publication status and date: Edited (no change to conclusions), published in Issue 1, 2020.

Citation: Regnaux JP, Davergne T, Palazzo C, Roren A, Rannou F, Boutron I, Lefevre-Colau MM. Exercise programmes for ankylosing spondylitis. *Cochrane Database of Systematic Reviews* 2019, Issue 10. Art. No.: CD011321. DOI: 10.1002/14651858.CD011321.pub2.

ABSTRACT

Background

Exercise programmes are often recommended for managing ankylosing spondylitis (AS), to reduce pain and improve or maintain functional capacity.

Objectives

To assess the benefits and harms of exercise programmes for people with AS.

Search methods

We searched CENTRAL, the Cochrane Library, MEDLINE Ovid, EMBASE Ovid, CINAHL EBSCO, PEDro, Scopus, and two trials registers to December 2018. We searched reference lists of identified systematic reviews and included studies, handsearched recent relevant conference proceedings, and contacted experts in the field.

Selection criteria

We included reports of randomised controlled trials (RCT) of adults with AS that compared exercise therapy programmes with an inactive control (no intervention, waiting list) or usual care.

Data collection and analysis

We used standard Cochrane methodology.

Main results

We included 14 RCTs with 1579 participants with AS. Most participants were male (70%), the median age was 45 years (range 39 to 47), and the mean symptom duration was nine years. The most frequently used exercises were those designed to help improve strength, flexibility, stretching, and breathing. Most exercise programmes were delivered along with drug therapy or a biological agent. We judged most of the studies at unclear or high risk of bias for several domains. All 14 studies provided data obtained immediately upon completion of the exercise programme. The median exercise programme duration was 12 weeks (interquartile range (IQR) 8 to 16). Three studies (146 participants) provided data for medium-term follow-up (< 24 weeks after completion of the exercise programmes), and one

(63 participants) for long-term follow-up (> 24 weeks after completion of the exercise programmes). Nine studies compared exercise programmes to no intervention; five studies compared them to usual care (including physiotherapy, medication, or self-management).

Exercise programmes versus no intervention

All data were obtained immediately upon completion of the exercise programme.

For physical function, measured by a self-reporting questionnaire (the Bath Ankylosing Spondylitis Functional Index (BASFI) scale, 0 to 10; lower is better), moderate-quality evidence showed a no important clinically meaningful improvement with exercise programmes (mean difference (MD) -1.3, 95% confidence interval (CI) -1.7 to -0.9; 7 studies, 312 participants; absolute reduction 13%, 95% CI 17% to 9%).

For pain, measured on a visual analogue scale (VAS, 0 to 10, lower is better), low-quality evidence showed an important clinically meaningful reduction of pain with exercise (MD -2.1, 95% CI -3.6 to -0.6; 6 studies, 288 participants; absolute reduction 21%, 95% CI 36% to 6%).

For patient global assessment of disease activity, measured by a self-reporting questionnaire (the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) scale, 0 to 10, lower is better), moderate-quality evidence showed no important clinically meaningful reduction with exercise (MD -0.9, 95% CI -1.3 to -0.5; 6 studies, 262 participants; absolute reduction 9%, 95% CI 13% to 5%).

For spinal mobility, measured by a self-reporting questionnaire (the Bath Ankylosing Spondylitis Metrology Index (BASMI) scale, 0 to 10, lower is better), very low-quality evidence showed an improvement with exercise (MD -0.7 95%, -1.3 to -0.1; 5 studies, 232 participants) with no important clinical meaningful benefit (absolute reduction 7%, 95% CI 13% to 1%).

For fatigue, measured on a VAS (0 to 10, lower is better), very low-quality evidence showed a no important clinically meaningful reduction with exercise (MD -1.4, 95% CI -2.7 to -0.1; 2 studies, 72 participants; absolute reduction 14%, 95% CI 27% to 1%).

Exercise programmes versus usual care

All data were obtained immediately upon completion of the exercise programme.

For physical function, measured by the BASFI scale, moderate-quality evidence showed an improvement with exercise (MD -0.4, 95% CI -0.6 to -0.2; 5 studies, 1068 participants). There was no important clinical meaningful benefit (absolute reduction 4%, 95% CI 6% to 2%).

For pain, measured on a VAS (0 to 10, lower is better), moderate-quality evidence showed a reduction of pain with exercise (MD -0.5, 95% CI -0.9 to -0.1; 2 studies, 911 participants; absolute reduction 5%, 95% CI 9% to 1%). No important clinical meaningful benefit was found.

For patient global assessment of disease activity, measured by the BASDAI scale, low-quality evidence showed a reduction with exercise (MD -0.7, 95% CI -1.3 to -0.1; 5 studies, 1068 participants), but it was not clinically important (absolute reduction 7%, 95% CI 13% to 1%) with important clinical meaningful benefit

For spinal mobility, measured by the BASMI scale, very low-quality evidence found a no important clinically meaningful improvement with exercise (MD -1.2, 95% CI -2.8 to 0.5; 2 studies, 85 participants; absolute reduction 12%, 95% CI 5% less to 28% more). There was no important clinical meaningful benefit.

None of the studies measured fatigue.

Adverse effects

We found very low-quality evidence of the effect of exercise versus either no intervention, or usual care. We are uncertain of the potential for harm of exercises, due to low event rates, and a limited number of studies reporting events.

Authors' conclusions

We found moderate- to low-quality evidence that exercise programmes probably slightly improve function, may reduce pain, and probably slightly reduce global patient assessment of disease activity, when compared with no intervention, and measured upon completion of the programme. We found moderate- to low-quality evidence that exercise programmes probably have little or no effect on improving function or reducing pain, when compared with usual care, and may have little or no effect on reducing patient assessment of disease activity, when measured upon completion of the programmes. We are uncertain whether exercise programmes improve spinal mobility, reduce fatigue, or induce adverse effects.

PLAIN LANGUAGE SUMMARY

Benefits and harms of exercise programmes for people with ankylosing spondylitis

Review question

We reviewed the evidence for the benefits and harms of exercise programmes for people with ankylosing spondylitis (AS).

Background

Exercise programmes are often recommended for people with AS, to reduce pain, and improve joint mobility or function.

Study characteristics

We searched for randomised controlled trials (RCT) to December 2018. We found 14 reports (1579 participants). Studies were performed in nine different countries. Most participants were men, aged 39 to 47 years old, who had symptom from 9 to 18 years. Mostly, the programmes included exercises developed to improve strength, flexibility, stretching, and breathing, and were added to drug therapy or a biological agent.

Key results

All data were obtained immediately upon completion of the exercise programme.

Exercise programmes versus no intervention

Exercise probably slightly improves function (moderate-quality evidence), slightly reduces patient-reported disease activity (moderate-quality evidence), and may reduce pain (low-quality evidence). We are uncertain of the effect on spinal mobility and fatigue (very low-quality evidence).

Physical function was measured on a self-reporting questionnaire, the Bath Ankylosing Spondylitis Functional Index (BASFI) scale (0 to 10; lower means better function). People who did not exercise rated their function at 4.1 points; those who exercised rated it 1.3 points lower (13% absolute improvement).

Pain was measured on a visual analogue scale (VAS, 0 to 10; lower means less pain). People who did not exercise rated their pain at 6.2 points; those who exercised rated it 2.1 points lower (21% absolute improvement).

Patient global assessment of disease activity was measured on a self-reporting questionnaire, the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI, 0 to 10, lower means less disease activity). People who did not exercise rated their disease activity at 3.7 points; those who exercised rated it 0.9 points lower (9% absolute improvement).

Spinal mobility was measured on a self-reporting questionnaire, the Bath Ankylosing Spondylitis Metrology Index (BASMI, 0 to 10, lower means better mobility). People who did not exercise rated their spinal mobility at 3.8 points; those who exercised rated it 0.7 points lower (7% absolute improvement).

Fatigue was measured on a VAS (0 to 10, lower means less fatigue). People who did not exercise rated their fatigue at 3 points; those who exercised rated it 1.4 points lower (14% absolute improvement).

Exercise programmes versus usual care

Exercise probably results in little or no improved function or reduced pain (moderate-quality evidence), and may have little or no effect in reducing patient-reported disease activity (low-quality evidence). We are uncertain of the effect on spinal mobility (very low-quality evidence).

Physical function. People who received usual care rated their function at 3.7 points on the BASFI; those who exercised rated it 0.4 points lower (4% absolute improvement).

Pain. People who received usual care rated their pain at 3.7 points on a 10-point VAS; those who exercised rated it 0.5 points lower (5% absolute improvement).

Patient global assessment of disease activity. People who received usual care rated their disease activity at 3.7 points on the BASDAI; those who exercised rated it 0.7 points lower (7% absolute improvement).

Spinal mobility. People who received usual care rated their spinal mobility at 8.9 points on the BASMI; those who exercised rated it 1.2 points lower (12% absolute improvement).

None of the studies measured fatigue.

Adverse effects (AE)

One of 67 participants in the exercise groups, and none of 43 participants in the control groups, experienced an AE.

Quality of the evidence

We downgraded the evidence due to issues with study design, variability between interventions, and not enough data, resulting in a rating of moderate to very low-quality evidence across outcomes.

Summary of findings for the main comparison. Exercise programmes compared to no intervention for ankylosing spondylitis

Exercise programmes compared to no intervention						
Patient or population: adults with ankylosing spondylitis **Setting:** international hospitals, outpatient clinics, or home **Intervention:** exercise programmes **Comparison:** no intervention						
Outcomes	**Anticipated absolute effects* (95% CI)**		**Relative effect (95% CI)**	**№ of participants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
	Risk with no intervention	**Risk with exercise programmes**				
Physical function assessed with self-report questionnaire BASFI scale (0 (easy) to 10 (impossible)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean physical function in the control groups was 4.1[a]	The mean physical function in the exercise groups was 1.3 lower (1.7 lower to 0.9 lower)	-	312 (7 RCTs)	⊕⊕⊕⊖ MODERATE[b]	13% absolute reduction (95% CI 17% to 9%) 32% relative change (95% CI 23% to 42%) NNTB 3 (2 to 4)
Pain assessed with VAS scale (0 (no pain) to 10 (impossible)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean pain in the control groups was 6.2[a]	The mean pain in the exercise groups was 2.1 lower (3.6 lower to 0.6 lower)[c]	-	288 (6 RCTs)	⊕⊕⊖⊖ LOW[b,d]	MD -2.1 (95% CI -3.6 to -0.6) 21% absolute reduction (95% CI 36% to 6%) 34% relative change (95% CI 10% to 59%) NNTB = 3 (2 to 8)
Patient global assessment of disease activity assessed with self-report questionnaire BASDAI scale (0 (absent) to 10 (extreme)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean patient global assessment of disease activity in the control groups was 3.7[e]	The mean patient global assessment of disease activity in the exercise groups was 0.9 lower (1.3 lower to 0.5 lower)	-	262 (6 RCTs)	⊕⊕⊕⊖ MODERATE[b]	9% absolute reduction (95% CI 13% to 5%) 27% relative change (95% CI 15% to 39%) NNTB 4 (3 to 8)

Spinal mobility assessed with self-report questionnaire BASMI scale (0 (better) to 10 (very severe limitation)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean spinal mobility in the control groups was 3.8[e]	The mean spinal mobility in the exercise groups was 0.7 lower (1.3 lower to 0.1 lower)	-	232 (5 RCTs)	⊕⊖⊖⊖ VERY LOW [b, d, f]	7% absolute reduction (95% CI 13% to 1%) 18% relative reduction (95% CI 34% to 3%) NNTB 5 (3 to 14)
Fatigue assessed with VAS scale (0 (absent) to 10 (extreme)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean fatigue in the control groups was 3[e]	The mean fatigue in the exercise groups was 1.4 lower (2.7 lower to 0.1 lower)	-	72 (2 RCTs)	⊕⊖⊖⊖ VERY LOW [b,f,g]	14% absolute reduction (95% CI 27% to1%) 48% relative change (95% CI 5% to 91%) NNTB 3 (1 to 9)
Adverse effects associated with exercises Exercise programme duration: range 3 to 24 weeks	No adverse effects were reported in 43 control group participants	1 adverse effect was reported in 67 exercise group participants	Peto OR 6.25 (0.10 to 320.40)	110 (2 RCTs)[j]	⊕⊖⊖⊖ VERY LOW [g,h]	2% absolute increase (95% CI 5% less to 8% more) 152% relative change (95% CI 90% less to 5818% more) it was not possible to calculate NNTB as too few events were reported
Withdrawals because of adverse events Exercise programme duration: range 3 to 24 weeks	90 per 1000	96 per 1000 (68 to 134)	Peto OR 1.08 (0.74 to 1.57)	1343 (8 RCTs) [j]	⊕⊕⊖⊖ LOW [b, i]	1% absolute increase (95% CI 2% less to 4% more) 7% relative change (95% CI 23% less to 48% more) NNTB was not applicable as results were not statistically significant

*__The risk in the intervention groups__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: confidence interval; **OR:** odds ratio; **NNTB:** number needed to treat (benefit); **MD**: mean difference; **SMD**: standardized mean difference; **SD**: standard deviation; **BASFI:** Bath Ankylosing Spondylitis Functional Index; **VAS:** visual analogue scale

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[a] Souza 2017 is the source document for the control group baseline data
[b] Downgraded one level due to risk of detection bias for subjective outcomes (lack of blinding of participants)
[c] We calculated a pooled SMD and re-expressed it in MD, as the SMD multiplied by the control group baseline SD (SF-36 pain = 2.5 from Souza 2017)
[d] Downgraded one level for inconsistency; important heterogeneity
[e] Masiero 2011 is the source document for the control group baseline data
[f] Downgraded one level for imprecision; total number of participants less than 400 and large confidence intervals
[g] Downgraded one level for imprecision; low rate of events
[h] Downgraded two levels for risk of bias; no blinding, incomplete outcome reporting
[i] Downgraded one level for indirectness. Since only two studies explicitly monitored adverse events, we used dropouts or withdrawals for any reason as a major outcome measure to estimate adverse events
[i] Studies were included regardless of the comparator intervention

Summary of findings 2. Exercise programmes compared to usual care for ankylosing spondylitis

Exercise programmes compared to usual care

Patient or population: adults with ankylosing spondylitis
Setting: international hospitals, outpatient clinics, or home
Intervention: exercise programmes
Comparison: usual care (current practices included medication, self management, physiotherapy)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with usual care	Risk with exercise programmes				
Physical function assessed with self-report questionnaire BASFI scale (0 (easy) to 10 (impossible)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean physical function in the control groups was 3.7[a]	The mean physical function in the exercise groups was 0.4 lower (0.6 lower to 0.2 lower)	-	1068 (5 RCTs)	⊕⊕⊕⊖ MODERATE [b]	4% absolute reduction (95% CI 6% to 2%) 11% relative change (95% CI 5% to 16%) NNTB 10 (6 to 21)
Pain assessed with VAS scale (0 (no pain) to 10 (impossible)), at the end of intervention	The mean pain in the control groups was 3.7[a]	The mean pain in the exercise groups was 0.5 lower (0.9 lower to 0.1 lower) [c]	-	911 (2 RCTs)	⊕⊕⊕⊖ MODERATE [b]	MD -0.5 (95% CI -0.9 to -0.1) 5% absolute reduction (95% CI 9% to 1%)

Exercise programme duration: range 3 to 24 weeks						15% relative change (95% CI 2% to 22%) NNTB = 10 (7 to 68)
Patient global assessment of disease activity assessed with self-report questionnaire BASDAI scale (0 (absent) to 10 (extreme)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean patient global assessment of disease activity in the control groups was 3.7[a]	The mean patient global assessment of disease activity in the exercise groups was 0.7 lower (1.3 lower to 0.1 lower)	-	1068 (5 RCTs)	⊕⊕⊖⊖ LOW [a,d]	7% absolute reduction (95% CI 13% to 1%) 19% relative change (95% CI 3% to 35%) NNTB 6 (3 to 52)
Spinal mobility assessed with self-report questionnaire BASMI scale (0 (better) to 10 (very severe limitation)), at the end of intervention Exercise programme duration: range 3 to 24 weeks	The mean spinal mobility in the control groups was 8.9[e]	The mean spinal mobility in the exercise groups was 1.2 lower (2.8 lower to 0.5 higher)	-	85 (2 RCTs)	⊕⊖⊖⊖ VERY LOW [a,d, f]	12% absolute change (95% CI 5% less to 28% more) 13% relative change (95% CI 6% less to 32% more) NNTB = NA
Fatigue		see comment	-	(0 RCTs)	-	No included studies measured this outcome
Adverse effects associated with exercises Exercise programme duration: range 3 to 24 weeks	No adverse effects were reported in 43 control group participants	1 adverse effect was reported in 67 exercise group participants	Peto OR 6.25 (0.10 to 320.40)	110 (2 RCTs) [i]	⊕⊖⊖⊖ VERY LOW[g, h]	2% absolute increase (95% CI 5% less to 8% more) 152% relative change (95% CI 90% less to 5818% more) it was not possible to calculate NNTB as too few events were reported
Adverse events Exercise programme duration: range 3 to 24 weeks		see comment		cannot be estimate	⊕⊖⊖⊖ VERY LOW[g,h]	Adverse events could not be calculate as events were not monitored or reported

*__The risk in the intervention groups__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: confidence interval; **OR:** odds ratio; **NNTB:** number needed to treat (benefit); **MD**: mean difference; **SMD**: standardized mean difference;

SD: standard deviation; **BASFI:** Bath Ankylosing Spondylitis Functionnal Index; **VAS:** visual analogic scale.

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[a] Rodriguez-Lozano 2013 is the source document for the control group baseline data.
[b] Downgraded one level due to risk of detection bias for subjective outcomes (lack of blinding of participants)
[c] We calculated a pooled SMD and re-expressed it as a MD; we multiplied the SMD by the control group baseline SD (VAS pain = 3.0 from Rodriguez-Lozano 2013)
[d] Downgraded one level for inconsistency; important heterogeneity
[e] Altan 2012 is the source document for the control group baseline data
[f] Downgraded one level for imprecision; total number of participants less than 400, and large confidence intervals
[g] Downgraded one level for imprecision; low rate of events
[h] Downgraded two levels for risk of bias; no blinding, incomplete outcome reporting
[i] Studies were included regardless of the comparator intervention

BACKGROUND

Description of the condition

Ankylosing spondylitis (AS) is a chronic, inflammatory rheumatic disease that mainly affects the axial skeleton and sacroiliac joints, causing characteristic inflammatory back pain (Braun 2003; Braun 2010; van der Heijde 2008). The inflammatory back pain is due to sacroiliitis and spondylitis, and to the formation of syndesmophytes, leading to ankylosis of the spine (Baraliakos 2005). AS can start early, and often affects young adults. Men are more affected than women, with a ratio of 2:1 (Braun 2007). The disease affects about 0.1% to 1.4% of the population, depending on the geographical region (Boonen 2006), and is closely associated with positivity for human leukocyte antigen 27 (Dougados 2011). In a recent systematic review, the estimated AS prevalence was reported to be 18.6/10,000 in Europe, 18.0/10,000 in Asia, 12.2/10,000 in Latin America, 39.9/10,000 in North America, and 7.4/10,000 in Africa (Dean 2014). The number of AS cases is estimated to range from 1.30 million to 1.56 million in Europe and 4.63 million to 4.98 million in Asia. The incidence ranges from 0.5 to 14 per 100,000 people per year, depending on the country (Braun 2007).

The main clinical features of AS are back pain and reduced mobility, caused by inflammation in the axial skeleton spinal region. Approximately one-third of individuals report peripheral joint involvement, most often of the hip, shoulder, and knee joints. AS may also be associated with extra-articular manifestations, including enthesitis, anterior uveitis, inflammatory bowel disease, and inflammatory skin conditions (Braun 2007). Enthesitis (inflammation of the entheses, the sites at which tendons or ligaments insert into the bone) is typical of, and a key problem in, AS, and occurs at peripheral joints, generally the hip, shoulder, knee, or heel. AS may result in varying degrees of structural and functional impairments and reduced general health (Dagfinrud 2005). The severity of symptoms and radiographic progression of the disease vary considerably: longer disease duration, increasing age, and smoking are associated with decreased functioning (Boonen 2006). A cohort study found individuals with a high C-reactive protein (CRP) level and syndesmophytes to be at risk for radiographic progression of the spine (Poddubnyy 2012), an indicator of disease severity (Pradeep 2008). However, the major sequela of AS is decreased quality of life. Like many chronic diseases, AS is associated with high medical and socioeconomic costs: in a systematic review, Palla 2012 estimated that AS represents a total cost of USD 31.766 per year for individuals with increased functional disability and severe disease. About 20% of individuals with AS experience disability at work (Reveille 2012). Boonen 2006 found that AS had considerable impact on healthcare costs and workforce participation.

In AS, treatments are expected to reduce the pain and stiffness of the back and sacroiliac joints, and improve spine and peripheral joint mobility (Boonen 2004). Current recommendations for the global management of AS combine appropriate medication and exercises as the two cornerstones of treatment (Braun 2010). Pharmacological therapies have greatly improved disease management (Vliet 2009). Biologic therapies have been efficacious and have changed the management landscape of AS and axial spondyloarthritis (SpA (Baraliakos 2012)), particularly with the introduction of anti-tumour necrosis factor (TNF) agents (Vliet 2009). However, some individuals with AS (20% to 40%) do not respond well to pharmacological treatments (Dougados 2011). Whether these treatments can prevent structural change is unclear. Non-steroidal anti-inflammatory drugs (NSAIDs) seem to affect new bone formation, and some data suggest that they can positively affect the radiographic progression of axial SpA (Poddubnyy 2012). The benefits of biologic treatment on the structural progression of the disease are still debated (van der Heijde 2008). Recent data indicate that biologic therapy can slow the structural progression of AS (Haroon 2013).

Description of the intervention

Exercise programmes have been used to treat AS and remain a part of its management (Braun 2010). Up to 10% to 20% of individuals with AS receive physical therapy in the United States (Reveille 2012). According to the typical clinical features of AS, exercise programmes have focused on improving or maintaining spinal and thoracic mobility. Recently, studies have been designed to target other aspects of physical fitness, and to develop muscular strength and aerobic capacity (Giannotti 2014). A growing body of evidence suggests a dose–response relationship between exercise and health effects, as for drugs, so the effect of exercise depends on the individual's adherence to the prescribed programme (Arem 2015; Vidoni 2015).

How the intervention might work

Exercise programmes are associated with different hypothesised mechanisms of effect (Kujala 2009; Hagen 2012), and may benefit people with AS (Altan 2012). They may help avoid stiffness, and improve or maintain functional capacity by moving joints, especially during back stretching, posture control, muscle strengthening, pulmonary function, and cardiovascular fitness (Fernández-de-Las-Peñas 2005; Niedermann 2013). Other benefits include improving quality of life and reducing pain (Singh 2013). Different exercise programmes are available (Van den Berg 2012). Some clinical trials have reported that the use of tai chi, global posture re-education, exercises combined with spa treatments, or multimodal exercise programmes may be effective but the effect of different types of exercise programmes remains unclear (Wang 2009). The exercises are extremely heterogeneous: they can vary in dosage, type of exercise, components, modes, and settings (Slade 2016). The optimal mode of delivery, optimal frequency and duration of treatment, and in particular whether particular components of exercise modalities can improve the clinical outcome need to be explored. A Cochrane Review of 11 RCTs of individuals with AS concluded that exercises have a small but significant positive effect on pain, spinal mobility, physical function, and patient global assessment (Dagfinrud 2008).

Why it is important to do this review

Given the publication of new RCTs on exercise programmes in AS, a comprehensive systematic review is important to examine the evidence for exercise for people with AS.

OBJECTIVES

To assess the benefits and harms of exercise programmes on physical function, pain, fatigue, and global assessment of disease activity in people with ankylosing spondylitis.

METHODS

Criteria for considering studies for this review

Types of studies

We included RCTs.

Types of participants

We included studies involving adults (18 years or older, with no upper age limit) with a diagnosis of ankylosing spondylitis (AS) according to the modified New York criteria (Van der Linden 1984), with critical features of visible structural damage on the sacroiliac joint on X-rays.

We excluded individuals with non-radiographic axial spondyloarthritis (SpA (Slobodin 2015)), as defined by the European Spondyloarthropathy Study Group and the Amor criteria, or the Assessment of Spondyloarthritis International Society (ASAS) criteria for axial SpA (Rudwaleit 2009; Van den Berg 2013).

We included studies with other populations only if we were able to extract data for the AS group separately.

Types of interventions

We defined exercise as 'a form of physical activity that is planned, structured and repeated over a period of time' (Bouchard 2012), with the intention of 'reducing pain and disability and improving overall health' (Abenhaim 2000; Hayden 2012).

We included interventions that delivered any type of exercise. The exercises could aim to improve any combination of stretching, flexibility, mobilising, balance, aerobic, strengthening, or functional training. We considered multimodal physical therapy interventions if one group of participants received exercise as part of a multimodal intervention and the comparison group received a non-exercise intervention (attentional, control intervention), or no intervention.

We considered trials that included co-interventions. We included trials that compared an exercise programme plus a co-intervention versus the co-intervention alone (e.g. exercise training plus a non-steroidal anti-inflammatory drug (NSAID) versus the NSAID alone). The only difference between groups was the exercise intervention.

We included exercise programmes carried out in any setting or location (home, inpatient clinic, hospital, or elsewhere), with any type of delivery (individual, group, or mixed); they could be land-based or water-based.

We included specific programmes, such as tai chi or yoga.

We considered any trial comparing exercise programmes with:

- No exercise (attention, no treatment, waiting list control). Participants were asked not to practice exercises during the study period.
- Usual care (participants could practice exercises as usual).

We excluded trials with general activities (e.g. swimming or walking) that required only movements, and did not meet our definition of exercise.

Types of outcome measures

We assessed a core set of outcome measures recommended by the ASAS (www.asas-group.org; Sieper 2009; Van der Heijde 1997), and the 1999 conference on Outcome Measures for Rheumatoid Arthritis Clinical Trials (Van der Heijde 1999). We extracted all outcomes for analysis according to the following preferred hierarchy:

Major outcomes

Physical function

If data on more than one physical function scale were provided for a trial, we extracted data on the physical function scale that was highest on the following list:

- Physical function (Bath Ankylosing Spondylitis Functional Index (BASFI))
- Dougados Functional Index (DFI)
- Health Assessment Questionnaire for AS (HAQ-AS)

Pain

If data on more than one type of pain scale were provided for a trial, we extracted data on the type of pain scale that was highest on the following list, according to a previously described hierarchy of pain-related outcomes (Sieper 2009).

In a visual analogue scale (VAS) or numerical rating scale (NRS):

- Total back or spine pain (Bath Ankylosing Spondylitis Disease Activity Index (BASDAI))
- Overall pain
- Back or spine pain at night
- Overall pain at night

Patient global assessment of disease activity

If data on more than one patient global assessment of disease activity scale were provided for a trial, we extracted data on the patient global assessment of disease activity scale that was highest on the following list:

- BASDAI
- Patient global VAS or NRS (global disease activity in the previous week)
- Stiffness VAS or NRS (duration of morning stiffness, spine, last week

Spinal mobility

If data on more than one spinal mobility scale were provided for a trial, we extracted data on the spinal scale that was highest on the following list:

- Schober test score
- Lateral spinal flexion
- Cervical rotation
- Occiput to wall movement
- Chest expansion
- Bath Ankylosing Spondylitis Metrology Index (BASMI)

We considered including BASMI and other spinal scales as separate outcomes.

Fatigue

- BASDAI fatigue question.

Safety

- Withdrawals due to adverse events (AEs).
- Severe AE outcomes: inpatient hospitalisation, life-threatening events, or death
- Adverse effects associated with the exercise intervention: we extracted the proportion of participants who experienced adverse effects related to exercise programmes (including joint or muscle contractures, fatigue, pain, falls, functional limitations)

Minor outcomes

Quality of life

- Medical Outcomes Survey Short Form-36 (SF-36)
- Ankylosing Spondylitis Quality of Life Instrument (ASQoL)
- EuroQol (EQ-5D)

Acute-phase reactant

- C-reactive protein (CRP) level (mg/L) or erythrocyte sedimentation rate (ESR)

Physician global assessment

Peripheral joints, entheses (pain, swelling, and tenderness)

- Number of swollen joints (44-joint count (Braun 2007))
- Validated enthesitis score, such as the Maastricht Ankylosing Spondylitis Enthesis Score (MASES), the University of California, San Francisco Index, and the Berlin Index

Timing of outcome assessment

We extracted outcome measures at the following three times points:

- end of intervention – measured immediately after completion of the exercise programme
- medium-term follow-up – < 24 weeks after completion of the exercise programme
- long-term follow-up – ≥ 24 weeks after completion of exercise

Search methods for identification of studies

Electronic searches

We searched the following electronic databases for primary studies, from database inception up to the search date. The last search was in 14 December 2018:

- the Cochrane Central Register of Controlled Trials (CENTRAL; 2018, Issue 12) in the Cochrane Library (searched 14 December 2018);
- MEDLINE Ovid (1946 to 14 December 2018);
- Embase Ovid (1974 to 14 December 2018);
- CINAHL EBSCO (Cumulative Index to Nursing and Allied Health Literature; 1982 to 14 December 2018);
- PEDro (www.pedro.org.au/; searched 14 December 2018);
- Scopus (searched 14 December 2018).

We searched the *Cochrane Database of Systematic Reviews* (14 December 2018) and the Database of Abstracts of Reviews of Effect (up to 14 December 2018) to identify relevant systematic reviews.

The queries combined free text words and controlled vocabulary. The search strategy was based on synonyms of ("exercise") AND "spondyloarthritis". The Cochrane Musculoskeletal Review Group's Information Specialist helped to develop each search strategy.

The electronic search strategy for MEDLINE is outlined in Appendix 1. We adapted this search strategy for use with other databases. We used the 'optimal sensitive search strategies' designed to identify clinical trials, described by Lefebvre 2011.

We did not restrict the search by language of publication or publication status.

Searching other resources

We hand-searched the reference lists of selected trials and systematic reviews identified from electronic searches, and also searched in Google and Google Scholar.

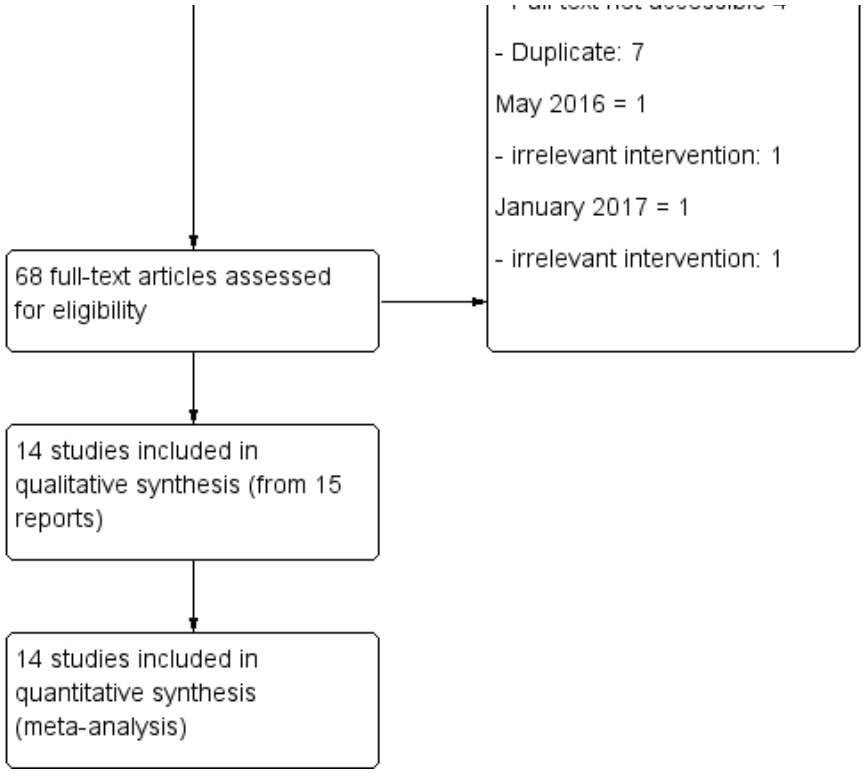
We searched the proceedings of the conferences of the American College of Rheumatology (on July 2013, November 2014) , European League Against Rheumatism (EULAR) (October 2013; November 2014), and Osteoarthritis Research Society International (on April 2013, April 2014) available online, and contacted authors and field experts for any additional published or unpublished data.

We searched the US National Institutes of Health Ongoing Trials Register ClinicalTrials.gov (www.clinicaltrials.gov; searched December 2018) and the World Health Organization International Clinical Trials Registry Platform (apps.who.int/trialsearch/; searched December 2018) to identify any studies in progress.

We present a flow diagram of search results and selection of studies in Figure 1.

Figure 1. Study flow diagram. Search results from original June 2015 literature search, and May 2016 and January 2017 updates

Figure 1. (Continued)



Data collection and analysis

Selection of studies

We removed duplicate records from the references identified. Two review authors (JPR, TD) independently reviewed the titles and abstracts of citations identified from the search strategy to select potentially relevant studies. Then, we obtained the full text of all potentially eligible studies and screened them for inclusion, according to the eligibility criteria. We resolved disagreements by reaching a consensus, or by consulting a third review author (MMLC) if necessary. We linked multiple reports relating to the same trial, or to trials with potentially overlapping populations. If the possibility of overlapping populations could not be excluded, we selected the most recent trial.

Data extraction and management

Two review authors (TD, JPR, or MMLC) independently extracted the results of individual trials by using a standardised, piloted extraction form, accompanied by a codebook. Disagreements were resolved by reaching consensus, or by consulting a third review author if necessary. The extraction form, based on other forms used by the Cochrane Musculoskeletal Review Group, was pilot tested with five reports of RCTs.

We extracted the following information:

1. Trial characteristics (funding, settings and number of centres, country, study design);
2. Participant characteristics (age, sex, measure of functional status, level of pain, description of radiographic damage, biologic medications, NSAIDs, corticosteroids or other drugs, coexisting diseases, other);
3. Intervention characteristics:
   a. number of intervention groups;
   b. content and type of each intervention (details);
   c. qualitative data: a detailed description of the interventions, including the different components of the programme received by each group, mode of delivery (individual, group, over internet), with or without supervision (face-to-face or at home), clinical expertise and background of the healthcare professionals who provided the exercise programmes (physiotherapist, fitness instructor, registered nurse, other), and adherence. We followed the reporting of Saunders 2016 to evaluate adherence by including: (1) attendance at the exercise programme sessions, and (2) compliance with the protocol or exercise instructions during the training sessions.
   d. quantitative data: the number of sessions, timing and duration of each session, duration of each component, and overall duration intensity. We collected these data as more frequent interventions, conducted over a long time, may influence outcomes.
4. Outcomes reported, including individual effect measures used (continuous or dichotomous data) and timing of outcome measurement.
5. AEs: we reported any AEs and/ or adverse effects related to the interventions in each group.
6. Economic data: we summarised economic evaluations in additional tables when available.

When necessary, we used PlotDigitizer to approximate data from graphs (arohatgi.info/WebPlotDigitizer/index.html). We entered the data into Review Manager 5 (RevMan 2014), and checked for accuracy.

Assessment of risk of bias in included studies

We evaluated the risk of bias in each included study according to Cochrane's 'Risk of bias' tool (Higgins 2011a). Two review authors (TD, JPR, or MMLC) independently examined seven specific domains: sequence generation, allocation concealment, blinding of participants, blinding of personnel who delivered exercise programmes, blinding of outcome assessors, incomplete outcome data, and selective outcome, plus other potential sources of bias (i.e. imbalanced baseline characteristics, small study participants, lack of power calculation, no assessments of attendance).

We separately assessed the blinding of self-reported subjective outcomes (e.g. pain, function, health-related quality of life) and the blinding of independent outcome assessors to objective outcomes (such as AEs).

Studies were classified at low risk of bias if all domains were assessed at low risk for potential bias; high risk of bias if one or more categories was assessed at high risk of bias; and unclear risk of bias if one or more key domains was assessed at unclear risk of bias. We resolved disagreements by discussion, or by consulting a third review author if necessary.

Measures of treatment effect

We calculated point estimates and 95% confidence intervals (CIs) for outcomes of individual RCTs whenever possible.

We summarised the intervention effect estimates in a meta-analysis only when estimates displayed sufficient clinical and statistical homogeneity. The estimate of the common treatment effect was the weighted average of the individual estimates for each study.

If the meta-analysis resulted in statistically significant overall estimates, we transformed these treatment effect measures (pooled estimate of the relative risk or SMD) into measures that are clinically useful in daily practice, such as number needed to treat for an additional beneficial outcome (NNTB), or number needed to treat for an additional harmful outcome (NNTH), and the absolute or relative improvement in the original units. We calculated the absolute risk difference and relative percentage change by using the recommendations provided by the Musculoskeletal Review Group (musculoskeletal.cochrane.org).

We assumed a minimal clinically important difference (MCID) of 1.5 on a 10-cm scale for pain, patient global assessment of disease activity, physical function, or physician global assessment. We defined an important clinical benefit as an outcome improvement that was more than 15% for an absolute change, and more than 20% for a relative change (Tubach 2012). We did not consider outcome changes that were below these values to be clinically important.

For dichotomous data

We analysed AEs by using Peto odds ratios (Peto ORs).

For dichotomous outcomes, we calculated the NNTB or NNTH from the control group event rate (unless the population event rate was known) and the relative risk, by using the Visual Rx NNT calculator (Cates 2008). We used the baseline values observed in the comparator group in the trials.

For continuous data

We summarised results, such as mean differences (MD), if the same tool was used to measure the same outcome across studies. We calculated the standardised MD (SMD) when the same outcome was measured with different units and methods of assessment across studies (e.g. pain scales). SMDs are calculated by dividing the MD by the standard deviation (SD); we calculated 95% CIs.

To enhance interpretability of continuous outcomes, we back-transformed pooled SMDs for overall pain and disability to an original 0 to 10 VAS for pain. When the direction of a scale (i.e. SF-36, 100 representing more favourable state of health) differed from the VAS for pain (10 defining high pain), we subtracted the mean from the maximum possible value for the scale, following the procedure recommended by Cochrane, and described in Chapter 9 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2011).

For continuous outcomes, we calculated the absolute risk difference as the mean difference between intervention and control groups in the original measurement units (divided by the scale), expressed as a percentage; the relative difference was calculated as the absolute change (or MD) divided by the baseline mean of the control group obtained from a representative trial, or the pooled baseline mean calculated in RevMan 5 by using the generic inverse variance method (Buchbinder 2015). We re-expressed outcomes pooled using SMDs as changes by multiplying by a representative control group baseline SD. We calculated the NNTB by using the Wells calculator software available at the Cochrane Musculoskeletal Review Group editorial office.

If we could not summarise results as described above, we reported them as 'other data' in narrative form, but did not include them in the meta-analysis (Deeks 2011).

Unit of analysis issues

For studies containing more than two intervention groups, we combined groups to create a single pair-wise comparison following the procedure recommended by Cochrane (Higgins 2011b).

Dealing with missing data

We contacted the original investigators to request any missing outcome data. If we did not receive a response, we sent two e-mail reminders, with two-week intervals.

For continuous outcomes with no SD reported, we calculated SDs from standard errors (SEs), 95% CIs, or P values (Higgins 2011c).

Assessment of heterogeneity

As recommended in the Cochrane Handbook for Systematic Reviews of Interventions (Deeks 2019, ch10.10), we assessed the presence of heterogeneity. We used the I^2 statistic: the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (Higgins 2011a). We interpreted the value of the I^2 statistic according to the following thresholds:

- 0% to 40%: might not be important
- 30% to 60%: may represent moderate heterogeneity
- 50% to 90%: substantial heterogeneity
- 75% to 100%: considerable heterogeneity (Deeks 2011).

We also computed the 95% CI for the I^2 statistic (Ioannidis 2007a), and the between-study variance Tau^2, estimated from the random-effects model (Rucker 2008). .

When we found substantial to considerable heterogeneity (severe heterogeneity), we checked the extracted data and insured that the numbers were correctly entered in the analysis software. When the number of trials was sufficient, we discussed the potential sources of heterogeneity by identifying a study that could be responsible of the presence of heterogeneity. As it is recommended, we did not exclude any study from the meta-analyses unless it can be considered as an outlier for an obvious reason (ie conflicting data). We also used the random-effects model with the DerSimonian and Laird approach to take into account the clinical differences between the studies included (Deeks 2011, Deeks 2019).

Assessment of reporting biases

To assess the presence of small study effects, we had planned to visually inspected funnel plots for each meta-analysis when the required statistical conditions were met (≥ 10 studies, no significant heterogeneity, and a ratio of the maximal to minimal variance across studies > 4).

Data synthesis

We performed a meta-analysis if the data from the studies were sufficiently clinically and statistically homogeneous. Because of large clinical heterogeneity between exercise programmes, participants, and characteristics, we used the random-model effects for all meta-analyses . We analysed and presented data separately by common control group intervention: exercise programmes versus no intervention, and exercise programmes versus usual care.

We analysed data at study completion, medium-term follow-up (< 24 weeks after study completion), and long-term follow-up (> 24 weeks after study completion).

In this review, we included studies with different characteristics, used different types of interventions, and reported effects on different outcomes measures. For a better description and standardisation, we presented a synthesis of these different characteristics in additional tables. We systematically described the key exercise programme components, according to the items recently proposed by Slade 2016 in the 'Characteristics of included studies' tables.

Subgroup analysis and investigation of heterogeneity

We planned to separate the data analysis on the basis of the control group intervention. We did not perform the other planned subgroup analyses (see the "Differences between protocol and review" section) because of the small number of studies in each group.

Sensitivity analysis

We did not perform any sensitivity analyses (see the "Differences between protocol and review" section).

'Summary of findings' tables

We included 'Summary of findings' tables to provide key information concerning the quality of evidence, the magnitude of effect of the interventions examined, and the sum of available data on the main outcomes, as recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2011a). We assessed the overall quality of the evidence for each main outcome by using the GRADE approach (Schünemann 2011b).

We developed 'Summary of findings' tables using GRADEpro GDT (GRADEpro GDT).

For the 'Summary of findings' tables, we included the following outcomes for each main comparison:

- Physical function
- Pain
- Patient global assessment of disease activity
- Spinal mobility
- Fatigue
- Adverse effects associated with exercise
- Adverse events

RESULTS

Description of studies

Results of the search

January 2015: we identified 806 citations after removing duplicates, and excluded 745 studies after screening titles and abstracts. In total, we selected 64 full-text reports for evaluation. After assessing all records, we included 11 unique studies. Among the 64 full-text reports, we contacted 18 authors (see Table 1): nine responded, and we obtained data for one study (Dönmez 2014). We identified two congress reports, but had insufficient information to include or extract the data, so we listed one study in the Characteristics of studies awaiting classification section. In addition, we identified four ongoing trials (Gallinaro 2016; Souza 2017; ChiCTR-TRC-14004650; NCT02098694). See Characteristics of ongoing studies.

Updated search in May 2016: we searched the listed electronic databases for reports of randomised controlled trials (RCT) published from January 2015 to May 2016. The search resulted in 75 records to screen. We assessed two full-text records to determine their eligibility. We included one new study (Garcia 2015).

Updated search in January 2017: we searched the listed electronic databases for RCT reports published from May 2016 to 31 January 2017. The search identified 51 records. We included two new studies identified in a previous search as ongoing trials (Gallinaro 2016; Souza 2017). Souza 2017 had published their data in a scientific journal, and Gallinaro 2016 had limited data published on ClinicalTrials.gov, and additional data on a thesis online website (www.teses.usp.br/teses/disponiveis/5/5169/tde-04112016-150051/fr.php).

A flow chart shows the overall search process in Figure 1.

We performed a further search in December 2018. We added those results to Characteristics of studies awaiting classification, and will incorporate them into the review at the next update.

Included studies

We provided a full description of each included study in the 'Characteristics of included studies' table. We also provided a descriptive summary of the information on trials, participants, and exercise programmes in additional tables (Table 2; Table 3; Table 4).

We included a total of 14 reports of RCTs. Reports were published between 1990 and 2017. Three trials were conducted in Turkey (Altan 2012; Dönmez 2014; Ince 2006), two in Spain (Garcia 2015; Rodriguez-Lozano 2013), two in Norway (Kjeken 2013; Sveaas 2014), two in Brazil (Gallinaro 2016; Souza 2017), and one in Canada (Kraag 1990), South Korea (Lim 2005), Italy (Masiero 2011), United Kingdom (Sweeney 2002), and Sweden (Widberg 2009).

Design

All included studies were RCTs, with a parallel-group design. There were no cross-over trials. Eleven studies included two groups, and three included three groups (Dönmez 2014; Gallinaro 2016; Masiero 2011). Most studies (N = 11, 79%) included fewer than 100 participants per group. The median number of participants per group was 26 (interquartile range (IQR): 15 to 29). All studies reported final values or pre–post differences for the exercise and control groups. We calculated individual study effects from means and standard deviations (SD). In one study, Masiero 2011 reported medians and IQRs. We used the formulas described by Hozo 2005 to estimate the mean and SD.

Participants

Participants were recruited from hospital departments (Gallinaro 2016; Ince 2006; Kjeken 2013; Lim 2005; Masiero 2011; Rodriguez-Lozano 2013; Souza 2017; Sveaas 2014; Sweeney 2002; Widberg 2009), clinics (Altan 2012), and arthritis patient associations (Garcia 2015; Kraag 1990; Sweeney 2002); the source was unclear in Dönmez 2014.

The 14 studies included a total of 1579 participants. The median sample size was 55 (range 35 to 73). The median age was 45 years (range 39 years to 47 years). Most participants were male (median 70% men). The modified New York criteria for ankylosing spondylitis (AS) diagnosis were most frequently used (71%). The median disease duration was nine years from diagnosis (range 9 years to 18 years). Many participants received non-steroidal anti-inflammatory drugs (NSAID (75%)); others received tumour necrosis factor (TNF) blockers (29%), or sulphasalazine (22%).

Interventions and comparators

Descriptions are provided in Table 4 and the 'Characteristics of included studies' tables.

The median exercise programme duration was 12 weeks (IQR 8 weeks to 16 weeks), with a median of three sessions (range two to seven) per week, and a median duration of 60 minutes per session (IQR 50minutes to 60 minutes). The description of dose components of exercise programmes was limited in three studies, in which exercise programmes were practiced at home (Kraag 1990; Lim 2005; Sweeney 2002). Intensity was variable and incompletely reported across studies.

Exercise programmes

For the 14 included studies, nine (64%) investigated exercise programmes alone in the experimental group (monomodal), and five (36%) combined exercise programmes with other interventions (education, self-management). The exercise programmes included a variety of components. The most commonly used components were strengthening exercises (64%), flexibility or stretching exercises (57%), and breathing exercises (50%). Most of the studies were land-based (11 studies). Two studies included an aquatic component in their exercise programmes (Garcia 2015; Kjeken 2013). One study was conducted only in water (Garcia 2015).

Exercise programmes were performed under the supervision of a therapist in nine studies (Altan 2012; Dönmez 2014; Gallinaro 2016; Garcia 2015; Ince 2006; Kraag 1990; Souza 2017; Sveaas 2014; Widberg 2009). Two studies instructed participants to undertake unsupervised exercise at home (Lim 2005; Sweeney 2002); three did not clearly report exercise supervision (Kjeken 2013; Masiero 2011; Rodriguez-Lozano 2013).

Nine studies reported the setting of the intervention. Six studies delivered exercise programmes in facilities (Garcia 2015; Ince 2006; Sveaas 2014; Widberg 2009), or combined them with home delivery (Masiero 2011; Rodriguez-Lozano 2013). Three studies were performed at participants' homes (Kraag 1990; Lim 2005; Sweeney 2002); five did not clearly mention where the exercise programmes were performed (Altan 2012; Dönmez 2014; Gallinaro 2016; Kjeken 2013; Souza 2017).

Control group interventions

Five included studies (36%) compared an exercise programme to usual care (Altan 2012; Kjeken 2013; Rodriguez-Lozano 2013; Sweeney 2002; Widberg 2009). Nine studies (64%) compared an exercise programme to no intervention (Dönmez 2014; Gallinaro 2016; Garcia 2015; Ince 2006; Kraag 1990; Lim 2005; Masiero 2011; Souza 2017; Sveaas 2014). For two of the nine studies, the description of the control intervention was unclear, and we had to contact the trial authors for additional information (Dönmez 2014; Ince 2006; Table 1). Based on the response from the two trial authors, we classified the control intervention as 'no intervention'.

Adherence to exercise programmes

We were unable to analyse the attendance, since attendance or compliance was not clearly reported in most of the included studies.

- Compliance: only four studies reported information, and reported that compliance was high. No data were provided (Altan 2012; Masiero 2011; Rodriguez-Lozano 2013; Widberg 2009).
- Two studies reported on attendance. Participants participated in at least 80% of the sessions in Sveaas 2014, and less than 50% of the exercise programmes in Gallinaro 2016.

Outcomes

The outcomes measured in each trial are summarised in Table 5, Table 6 and Table 7. For all 14 studies, the end of the intervention was considered the final data collection point (range 3 to 24 weeks).

Major Outcomes

Among the main outcomes (Table 5, Table 6), most trials included a measure of physical function (Bath Ankylosing Spondylitis Functional Index (BASFI), N = 12), and global patient assessment of disease activity (Bath Ankylosing Spondylitis Disease Activity Index

(BASDAI), N = 11); fewer included measures of overall pain (N = 9), fatigue (N = 2), or adverse effects (N = 2). For spinal mobility, the Bath Ankylosing Spondylitis Metrology Index (BASMI) was the most commonly reported (N = 8), but other descriptors were also reported (chest expansion N = 6; distance occiput to wall distance N = 2; distance finger to floor N = 4; or the Schober test N = 3). No study explicitly reported adverse events. Only two studies monitored and reported adverse effects associated with the exercise intervention.

Minor outcomes

Quality of life was reported for five studies (Table 7): two studies used the Ankylosing Spondylitis Quality of Life (ASQoL) scale (Altan 2012; Rodriguez-Lozano 2013), three used the SF-36 (Dönmez 2014; Kjeken 2013; Souza 2017), and one used the SF-12 (Garcia 2015). Only Sveaas 2014 and Souza 2017 reported C-reactive protein (CRP) levels and erythrocyte sedimentation rates (ESR). No study reported peripheral joint modification scales.

Follow-up

Three studies reported data at medium-term follow-up, from 12 to 24 weeks (Altan 2012; Dönmez 2014; Masiero 2011). The mean duration follow-up period was 18 weeks. One study reported a 48-week long-term follow-up (Kjeken 2013). We contacted 10 trial authors requesting missing data for unreported or partially reported outcomes (Table 1).

Excluded studies

We excluded 54 studies at full-paper review, as described in (Figure 1). We excluded eight studies (Characteristics of excluded studies: ; Ciprian 2013; Colina 2009; Durmus 2009; Gunay 2012; Karahan 2016; Kraag 1990, Lee 2008; Masiero 2015); five were controlled but not randomised trials (Colina 2009; Durmus 2009; Gunay 2012, Lee 2008; Masiero 2015), one study was a duplicate of an included study (Kraag 1990); The intervention was irrelevant in two studies (Ciprian 2013; Karahan 2016).

Ongoing studies

See Characteristics of ongoing studies.

We identified two ongoing studies registered in the WHO ICTRP as potentially eligible, but results were not available. The two studies compared exercise programmes in Norway (NCT02098694), and China (ChiCTR-TRC-14004650).

Awaiting Studies

See Characteristics of studies awaiting classification.

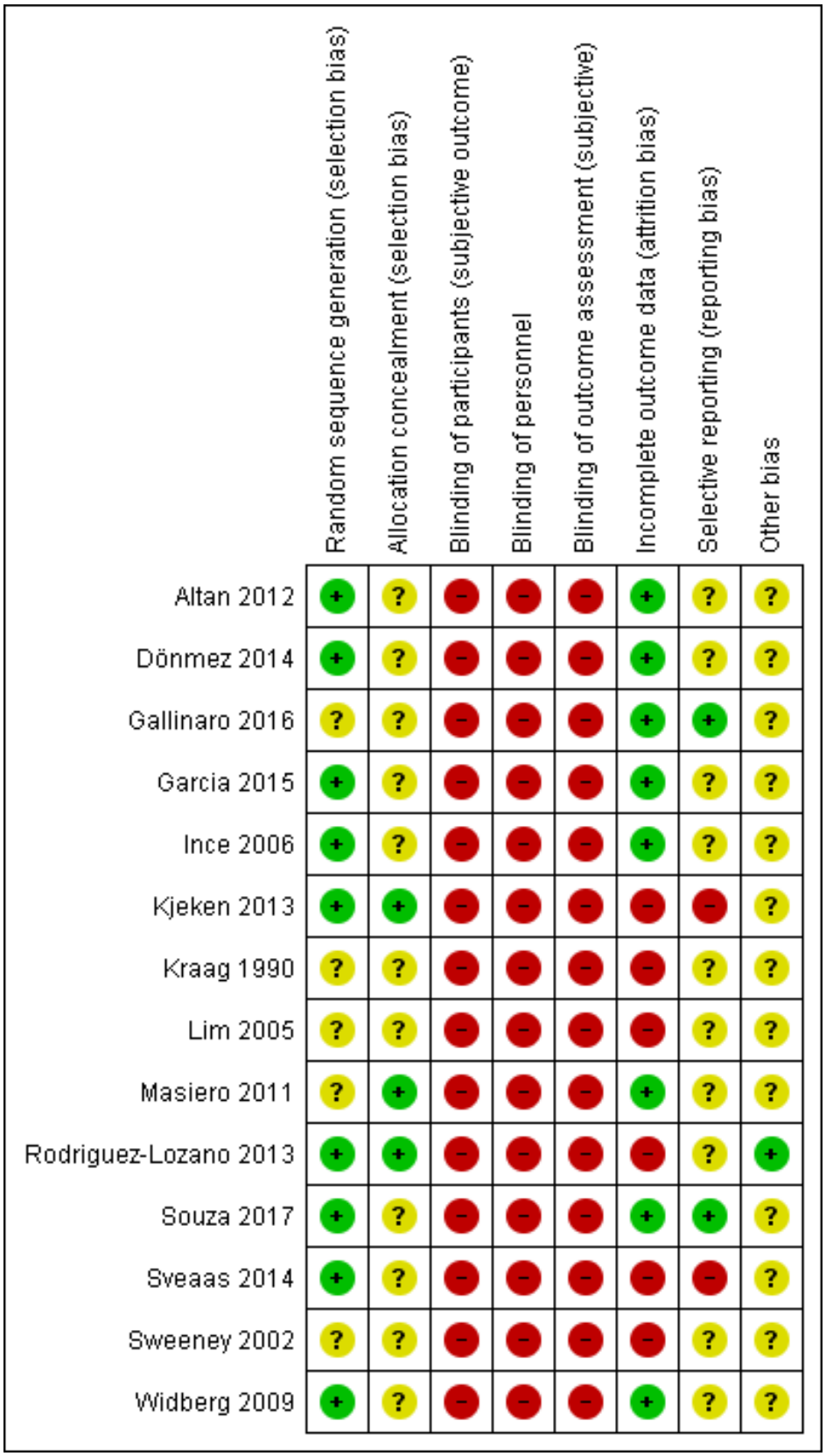
We identified one study as potentially eligible after we read the abstract, but we could not access the full-text article (Mesquita 2014). We tried to contact the trial authors for additional information, but received no response (Table 1).

We added one study report from our updated January 2018 search (Sveeas 2018), as we were unable to determine if the results of this study were new, or if it was a secondary analysis from the previous study (Sveaas 2014). We attempted to contact the authors, but received no response.

Risk of bias in included studies

The overall risk of bias assessment of the included studies is presented in Figure 2.

Figure 2. 'Risk of bias' summary: review authors' judgements about each risk of bias item for each included study



Allocation

Random sequence

We judged nine studies (64%) at low risk of bias, because they used and reported an appropriate method of randomisation (Altan 2012; Dönmez 2014; Garcia 2015; Ince 2006; Kjeken 2013; Rodriguez-Lozano 2013; Souza 2017; Sveaas 2014; Widberg 2009).

We assessed five trials (36%) at unclear risk of bias because the methods used to generate allocation sequence were not described, or were unclear (Gallinaro 2016; Kraag 1990; Lim 2005; Masiero 2011; Sweeney 2002).

Allocation concealment

We judged three studies (21%) at low risk of bias, since they provided adequate information on the method of allocation concealment (Kjeken 2013; Masiero 2011; Rodriguez-Lozano 2013).

For 11 studies (79%), the method used to conceal allocation sequence was unclear, or not described (Altan 2012; Dönmez 2014; Gallinaro 2016; Garcia 2015; Ince 2006; Kraag 1990; Lim 2005; Souza 2017; Sveaas 2014; Sweeney 2002; Widberg 2009).

Blinding

Participant and care provider blinding

We judged all studies at high risk of bias.

Blinding participants and care providers is difficult because of the nature of the intervention. Most of the included studies did not report information on blinding, or a masking procedure for treatment allocation or delivery. No studies reported using a blinding procedure (sham or attentional comparator, or blinding of study hypothesis (Boutron 2007)).

Outcome assessor

We judged all studies at high risk of bias. Most included studies used subjective outcomes (self-reporting, self-performance). Because participants were not blinded to treatment allocation, we considered the outcome assessors to be unblinded.

For studies that reported spinal mobility outcome, we considered them to be at unclear risk of bias, because it was impossible to evaluate whether assessors were blinded to treatment allocation (Gallinaro 2016; Ince 2006; Kraag 1990).

Incomplete outcome data

Eight studies (57%) reported no withdrawals, and drop-out rates were less than 20% at study completion (Altan 2012; Dönmez 2014; Gallinaro 2016; Garcia 2015; Ince 2006; Masiero 2011; Souza 2017; Widberg 2009). We judged these studies at low risk of bias.

Five studies (36%) reported higher rates (Kjeken 2013; Lim 2005; Rodriguez-Lozano 2013; Sveaas 2014; Sweeney 2002), and one trial reported an unbalanced rate between groups (Kraag 1990). Consequently, we judged these studies at high risk of bias. Only one study used an intention-to-treat approach for analysis (Souza 2017).

Selective reporting

Three studies (21%) had a registered protocol (Gallinaro 2016; Souza 2017; Sveaas 2014). We assessed two of them (14%) at low risk of reporting bias, because all outcomes reported were pre-specified in the protocol (Gallinaro 2016; Souza 2017).

We judged two studies (14%) at high risk of bias, because we found outcomes listed and not reported in the results section of the published report (Kjeken 2013, Sveaas 2014).

We judged the 10 remaining studies (71%) at unclear risk of reporting bias, because we could not compare the pre-specified outcomes with the reported ones.

Other potential sources of bias

We judged one study at low risk of bias because we identified no other potential source of bias (Rodriguez-Lozano 2013). Three studies (21%) reported a power sample calculation (Rodriguez-Lozano 2013; Souza 2017; Sveaas 2014).

Effects of interventions

See: **Summary of findings for the main comparison** Exercise programmes compared to no intervention for ankylosing spondylitis; **Summary of findings 2** Exercise programmes compared to usual care for ankylosing spondylitis

Exercise programmes versus no intervention

Major outcomes

Data were obtained at the end of the intervention; see Summary of findings for the main comparison.

Physical function (BASFI, 0 to 10 scale; lower score indicates higher function)

Seven studies (312 participants) found a reduction in physical function score with exercise versus no intervention at the end of the intervention (mean difference (MD) -1.3, 95% confidence interval (CI) -1.7 to -0.9); absolute risk difference 13% (95% CI 9% to 17%); relative change 32% (95% CI 23% to 42%); Analysis 1.1). The statistical heterogeneity was not important (I^2= 23%) . There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one point for high risk of bias; we rated the quality of the evidence as moderate (Dönmez 2014; Gallinaro 2016; Garcia 2015; Lim 2005; Masiero 2011; Souza 2017; Sveaas 2014).

Two studies (93 participants) found a reduction in physical function score with exercise at medium-term follow-up (overall 14 weeks (MD -1.5, 95% CI -1.8 to -1.2; Analysis 1.1)), which was clinically important (absolute risk difference 15% (95% CI 12% to 18%); relative change 57% (95% CI 44% to 67%)) (Dönmez 2014; Masiero 2011). The statistical heterogeneity was not important (I^2 = 0%).

Pain (VAS, 0 to 10; lower score indicates less pain)

The pooled analysis of six studies (288 participants) showed a decrease in pain with exercise at the end of the intervention (standardised mean difference (SMD) -0.82, 95% CI -1.4 to -0.25; Analysis 1.2; need to report the back-translated mean difference too here, as per methods (MD -2.1, 95% CI -3.6 to -0.6; 6 studies; absolute reduction 21%, 95% CI 36% to 6%) absolute reduction 21% (95% CI 6% to 3 6% better); relative reduction 34% (95% CI 10% to 59% better); (Dönmez 2014; Gallinaro 2016; Garcia 2015; Lim 2005; Masiero 2011; Souza 2017)). There was an important clinical meaningful benefit. The statistical heterogeneity was considerable

(I^2 = 81%). No rationale could be found to explain the observed severe heterogeneity. Because of study limitations, we downgraded the evidence by one level each for high risk of bias and imprecision; we rated the quality of the evidence as low. One study of 52 participants reported conflicting data (Kraag 1990) in their report. As the reported size effect (MD 0.4, 95% CI -0.2 to 0.9) was discordant and inconsistent with the findings of the other six studies, we decided not to include this study in the pooled analysis.

At medium-term follow-up (12 to 16 weeks), two studies (93 participants) assessed pain (Dönmez 2014; Masiero 2011) ; We found a statistically significant reduction of pain (SMD -2.46, 95% CI -5.19 to 0.28). The statistical heterogeneity was considerable (I^2 = 95%). No rationale could be found to explain the observed severe heterogeneity.

Patient global assessment of disease activity (BASDAI, 0 to 10 scale; lower score indicates lower disease activity)

Six studies (262 participants) found participants who exercised reported statistically significantly lower activity disease at the end of the intervention (MD -0.9, 95% CI -1.3 to -0.5; Analysis 1.3; absolute risk difference 9% (95% CI 5% to 13%); relative change 27% (95% CI 15% to 39%; Dönmez 2014; Gallinaro 2016; Garcia 2015; Masiero 2011; Souza 2017; Sveaas 2014)). The statistical heterogeneity was not important (I^2 = 18%). There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one level for high risk of bias; we rated the quality of the evidence as moderate.

Two studies (93 participants) found a statistically significant reduction in patient global assessment of disease activity with exercise at medium-term follow-up (MD -1.1, 95% CI -1.6 to -0.7; Analysis 1.3; Dönmez 2014; Masiero 2011). The statistical heterogeneity was not important (I^2 = 0%).

Spinal mobility

Schober test (tape distance in cm; longer distance indicates greater spinal mobility)

Three studies used the Schober test to assess spinal mobility. One study (51 participants) reported change from baseline, and found no evidence of difference between groups in spinal mobility (Kraag 1990). Two studies (85 participants) reported final values from a Schober test (Gallinaro 2016), and a modified Schober test (Ince 2006). Pooled results found no evidence of difference between groups (SMD 0.4, 95% CI -1.0 to 0.25) at the end of the intervention. The statistical heterogeneity was moderate (I^2 = 45%). There was no important clinical benefit. Because of study limitations, we downgraded the evidence by one level for high risk of bias, and by two levels for imprecision; we rated the quality of the evidence as very low.

BASMI (0 to 10 scale; lower score indicates greater spinal mobility)

Five studies (232 participants) found more spinal mobility with exercise at the end of the intervention (MD -0.7, 95% CI -1.3 to -0.1; Analysis 1.4; absolute risk difference 7% (95% CI 1% to 13%); relative change 18% (95% CI3% to 34%) (Dönmez 2014; Gallinaro 2016; Masiero 2011; Souza 2017; Sveaas 2014). The statistical heterogeneity was substantial (I^2 = 51%).There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one level each for high risk of bias, inconsistency, and imprecision; we rated the quality of the evidence as very low.

Two studies (93 participants) found more spinal mobility at medium-term follow-up (overall 14 weeks) with exercise (MD -1.4, 95% CI -2.0 to -0.8; Analysis 1.4; (Dönmez 2014; Masiero 2011)). The statistical heterogeneity was moderate (I^2 = 45%).

Fatigue (VAS, 0 to 10; lower score indicates less fatigue)

Two studies (72 participants) found a statistically significant reduction in fatigue with exercise versus no intervention at the end of the intervention (MD -1.4, 95 CI% -2.7 to -0.1; Analysis 1.5; absolute risk difference 14%, 95% CI 1% to 27%; relative change 48% (95% CI 5% to 91%; Garcia 2015; Masiero 2011). The statistical heterogeneity was substantial (I^2 = 70%). There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one level each for high risk of bias, imprecision, and inconsistency; we rated the quality of the evidence as very low.

At medium-term follow-up (24 weeks), one study (42 participants) found a reduction of fatigue with exercise (Masiero 2011). The mean fatigue with exercise was 2.1 on a 10-point VAS scale, with no exercise it was 3.7 (MD 1.6, 95% CI -2 to -1.2).

Minor outcomes

Quality of life (lower number is better)

We meta-analysed two of the five studies that assessed quality of life as an outcome.Two studies (85 participants) found inconclusive effects of exercise (MD 1.74, 95% CI -0.44 to 3.91; Analysis 1.6; Gallinaro 2016; Garcia 2015). The statistical Heterogeneity was not important (I^2 = 0%). For three studies, data were either not available (Dönmez 2014), or could not be extracted (Kjeken 2013; Souza 2017).

C-Reactive Protein (CRP) and Erythrocyte Sedimentation Rate (ESR)

Two studies (84 participants) reported data for CRP and ECR (Souza 2017; Sveaas 2014). For CRP, we found inconclusive results (MD 1.38, 95% CI -4.34 to 7.10 Analysis 1.7) at the end of the intervention . The statistical heterogeneity was substantial (I^2 = 71 %). No rationale could be found to explain the observed severe heterogeneity. For ESR, exercise reduced the level of ESR (MD -5.36, 95% CI -10.31 to -0.41 Analysis 1.8). The statistical heterogeneity was not important (I^2 = 0%).

Maastricht ankylosing spondylitis enthesitis score (MASES; 0 to 13 scale, lower is better)

One study (55 participants) reported final values at 16 weeks for the exercise and control groups on the 13-point MASES. No statistical difference was reported between groups (Gallinaro 2016).

Exercise programmes versus usual care

Major outcomes

Data were obtained at the end of the intervention; See Summary of findings 2.

Physical function (BASFI, 0 to 10 scale, lower score indicates higher function)

Five studies (1068 participants) found a reduction in physical function score (improvement; indicates higher function) with

exercise (MD -0.4, 95% CI -0.6 to -0.2; absolute risk difference 4%; 95% CI 2% to 6%; relative change 11%, 95% CI 5% to 16%; Analysis 2.1; Altan 2012; Kjeken 2013; Rodriguez-Lozano 2013; Sweeney 2002; Widberg 2009). The statistical heterogeneity was not important (I^2 = 0%). There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one level for high risk of bias; we rated the quality as moderate.

One study (53 participants) reported data at medium-term follow-up (Altan 2012). The results were inconclusive (MD -0.60, 95% CI -1.6 to 0.4; Analysis 2.1). One study (63 participants) reported data at long-term follow-up (48 weeks). The results were inconclusive (MD -0.10, 95% CI -0.84 to 0.64; Kjeken 2013).

Pain (VAS, 0 to 10; lower score indicates less pain)

Two studies (911 participants) reported pain, but used different scales (Rodriguez-Lozano 2013; Sweeney 2002). Rodriguez-Lozano 2013 used a VAS to measure pain; Sweeney 2002 used the Standford Self efficacy pain Scale (SES). Pooled analysis found a reduction in pain with exercise (SMD -0.2, 95% CI -0.3 to -0.03; Analysis 2.2; absolute reduction 6%, 95% CI 1% to 8% better; relative reduction 15%, 95% CI 2% to 22% better. The statistical heterogeneity was not important (I^2 = 0%). There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one level for high risk of bias; we rated the quality as moderate.

Patient global assessment of disease activity (BASDAI, 0 to 10 scale; lower score indicates lower disease activity)

Five studies (1068 participants) found a statistically significant reduction in patient global assessment of disease activity with exercise (MD -0.7, 95% CI -1.3 to -0.1; Analysis 2.3; absolute risk difference 7%, 95% CI 1% to 13%; relative change 19%, 95% CI 3% to 35%; Altan 2012; Kjeken 2013; Rodriguez-Lozano 2013; Sweeney 2002; Widberg 2009). The statistical heterogeneity was substantial (I^2 = 70%). No rationale could be found to explain the observed severe heterogeneity. There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one level each for high risk of bias and inconsistency; we rated the quality as low.

One study (93 participants) reported data at medium-term follow-up (24 weeks). A statistically significant improvement was found in patient global assessment of disease activity with exercise (MD -0.70, 95% CI -1.7 to 0.3; Altan 2012).

One study (63 participants) reported data at long-term follow-up (48 weeks); the results were inconclusive (MD -0.5, 95% CI -1.4 to 0.4; Kjeken 2013).

Spinal mobility

Schober test (tape distance in cm; longer distance indicates greater spinal mobility)

No study used the Schober test.

BASMI (0 to 10 scale; lower score indicates greater spinal mobility)

We meta analysed two of the three studies that used the BASMI (Altan 2012; Kjeken 2013; Widberg 2009). Kjeken 2013 did not report data. Two studies (85 participants) found inconclusive results for spinal mobility (MD -1.2, 95% CI -2.8 to 0.5; Analysis 2.4; absolute risk difference 12%. 95% CI 5% to 28%; relative change 163%, 95% CI 6% to 32%). The statistical heterogeneity was considerable (I^2 = 82%). No rationale could be found to explain the observed severe heterogeneity. There was no important clinical meaningful benefit. Because of study limitations, we downgraded the evidence by one level each for high risk of bias, inconsistency, and imprecision; we rated the quality as very low.

One study reported (53 participants) data at medium-term follow-up (24 weeks); the results were inconclusive for spinal mobility (MD -0.7 (95% CI -1.6 to 0.2; Analysis 2.4; Altan 2012).

One study (63 participants) reported data at long-term follow-up (48 weeks); the results were inconclusive (MD -0.00, 95% CI -0.6 to 0.6; Kjeken 2013).

Fatigue (VAS, 0 to 10; lower score indicates less fatigue)

We found no study measuring fatigue.

Minor outcomes

Quality of life (18-point ASQol scale; lower number is better)

Data from two studies (809 participants) found inconclusive evidence for quality of life (MD -0.36, 95% CI -1.68 to 0.95; Analysis 2.5; Altan 2012; Rodriguez-Lozano 2013). The statistical heterogeneity was moderate (I^2 = 46%)

C-Reactive Protein (CRP) and Erythrocyte Sedimentation Rate (ESR)

No study measured CRP or ESR.

MASES (0 to 13 scale, lower is better)

No study measured the enthesitis index.

Safety

Adverse effects (AE) associated with exercises

Two studies (110 participants) reported adverse effects related to exercises programmes (Altan 2012; Gallinaro 2016). Because of very low-quality evidence, we are uncertain of the effect of exercise programmes on AEs (Peto odds ratio (OR) 6.25, 95% CI 0.1 to 320; absolute risk difference 2%, 95% CI 5% fewer to 8% more; relative change 152%, 90% decrease to 5818% increase). The absolute numbers were very low: 1/67 in the exercise group versus 0/43 in the control group. Altan 2012 reported that one participant had an increase of back pain related to exercise. He did not mention whether this resulted in hospitalisation or not (Analysis 3.1). No adverse effects were considered serious. Because of study limitations, we downgraded the evidence by one point for high risk of bias and by two points for imprecision (large CI, small number of studies).

Adverse events

We found no study reporting adverse events.

Subgroup and sensitivity analyses

Given the small number of studies, we did not conduct subgroup analysis to explore the possible effect of type of delivery (supervision versus non-supervised, mono versus multimodal intervention) on estimated effect size. Neither did we conduct a sensitivity analysis, because we judged all studies at unclear or high risk of bias for most items.

Assessment of publication bias

We had planned to assess publication bias by visual inspection of funnel plots, but we did not generate funnel plots because of the limited number of studies (< 10), and the risk of an underpowered test. We were unable to determine the existence of publication bias.

DISCUSSION

Summary of main results

The main purpose of this review was to evaluate the benefit and harmful effects of exercise programmes for participants in trials of ankylosing spondylitis (AS). Overall, 14 randomised controlled trials (RCT; total of 1579 participants) met the inclusion criteria. Exercise programmes were examined alone in nine trials, and were combined with other interventions (education or self-management) in five trials. Exercise programmes were compared to usual care in five trials, and to no intervention (waiting list, advice, no exercise) in nine trials. Exercise programmes included different components, and were delivered in a variety of ways.

We found moderate- to low-quality evidence suggesting that exercise programmes, compared to no intervention, probably slightly improve function, may reduce pain (with an important clinical benefit), and probably slightly reduce patient assessment of disease activity at the end of the intervention. Whether there was an effect on spinal mobility and fatigue is unclear.

There is moderate- to low-quality evidence that compared with usual care (including physiotherapy, medication, or self-management), exercise programmes probably have little or no difference in improving function or reducing pain, and may have little or no difference in patient assessment of disease activity. We are uncertain whether exercise programmes improve spinal mobility.

All studies reported effects at the completion of the intervention. Only two studies assessed the medium-term follow-up effect of exercise programmes on physical function, patient global assessment of disease activity, and spinal mobility. One study reported long-term follow-up effects.

We have no clear evidence that exercise programmes can induce more adverse effects. Two studies reported adverse effects as an outcome. Only a small number of events were observed (one versus none in comparator groups). We were unable to draw any conclusions.

Overall completeness and applicability of evidence

The evidence provided by this review is limited to the 14 included RCTs that assessed the effects of exercise programmes versus no intervention, or usual care. We did not include three RCTs that were potentially eligible for this review, because their results have not yet been reported in full. Two were ongoing trials (ChiCTR-TRC-14004650; NCT02098694). According to the abstract of one trial, home exercises may improve function and spinal mobility, and ameliorate patient-assessment of disease activity after 10 weeks. However, the control intervention is unknown.

Whether the trial participants reflect individuals with AS undergoing treatment is difficult to determine. Most of the studies in this review included more than 70% men. Since AS seems to affect men and women differently, our results may have limited applicability to women (Dagfinrud 2005; Ramiro 2014). The median age of participants across the included studies was 45 years (interquartile range (IQR) 39 to 47), which is representative of the overall population of patients with AS (Ramiro 2014; Webers 2016). The mean duration of disease of the participants was nine years in the included studies.The effects of exercise programmes of this review should be extrapolated with cautious to people with a shorter disease duration. None of the studies investigated the impact of exercise programmes in people with early or newly diagnosed AS. Only one study included participants with a short disease duration (median 2.5 to 3.5 years), and showed beneficial effects of exercise on mobility (Widberg 2009). In the last decade, improved imaging techniques and criteria for early diagnosis of AS have facilitated earlier and more effective medical treatment (Liang 2015; Lubrano 2015). However, we lack studies of the efficacy of physical exercises on individuals with early forms of AS. Five studies included participants with a BASDAI score ≥ 3.5, corresponding to patients with a low patient-assessed disease activity (i.e. because the BASDAI scores were < 4/10 units). However, a cutoff of 3.9 to 4 is frequently used to define active disease, discriminating between people with well or poorly controlled disease, but this cutoff does not have a firm justification (Cohen 2006).

Most exercise programmes were delivered in conjunction with drug therapy (standard NSAIDs, disease-modifying anti-rheumatic drugs (DMARDs), or biological agents). The benefits of exercise programmes, depending on the type of drug therapy received, cannot be determined. Participants received different types of drug therapy. Nine of the included studies reported that 75% of the participants were taking NSAIDs. For seven studies, 29% of participants received anti-TNF agents. Four studies did not report or provide any information. No study specifically evaluated the efficacy of physical exercises with biologic versus standard NSAID or DMARD therapy. Since the introduction of TNF blockers, the role of physical exercise in individuals receiving TNF blockers has rarely been studied. We found only one RCT of participants receiving TNF blockers (Masiero 2011). In this study, for participants with clinically stabilised AS who had started TNF blocker therapy at least nine months previously, an educational-behavioural intervention and exercise training further improved spinal mobility, and reduced pain, stiffness, and disability. One hypothesis is that TNF blockers that reduce inflammation, pain, and fatigue may improve the efficacy of, and compliance with, regular physical exercises and activities, thereby resulting in better function and less disability (Maxwell 2015). Moreover, more motivated individuals are likely to spend longer periods of time on exercise, because they will have greater perceived benefits of exercise regimes (Dubey 2008). Further research should aim to determine the efficacy of exercise interventions in patients with AS receiving TNF blockers.

The included studies investigated a number of different types and combinations of exercise components. Breathing, strengthening, and stretching exercises were the most frequent exercise components. However, the components were incompletely described in most trials. For example the material used, who provided the intervention, how it was supervised, and where the exercise was delivered were often missing. However, the duration (mean 60 minutes) and frequency (three sessions/week) was similar among studies. The minimal effective dose and optimal level was unclear. The exercise dosage could not be explored with indirect statistical techniques, such as meta-regression. Thus, we

did not investigate heterogeneity by the type of exercise, because we were unable to isolate individual types of exercise from the programme reported by the authors.

Information was also lacking on adherence to exercise, which is important to assess with regular exercise and long disease duration. The optimal exercise programme for individuals with AS, and its efficacy, are still unknown. The poor reporting of non-pharmacological interventions is well known, despite the existence of reported guidelines (TIDieR), thereby limiting the implementation of research results in clinical practice (Hoffmann 2014). Slade 2016 recently developed a specific template, Consensus on Exercise Reporting Template (CERT), for reporting exercise programmes in clinical trials. We hope that this template will help authors improve the reporting of exercise in clinical trials.

Another issue relates to outcomes. We assessed important outcomes for participants based on the ASAS/EULAR recommendations, to show a short-term clinical benefit of exercise programmes versus no intervention (Sieper 2009; Van der Heijde 1997; www.asas-group.org). The most common outcomes measured in the included studies were physical function (N = 12, 86%), patient global assessment of disease activity (N = 11, 79%), spinal mobility and pain (N = 7, 50%). Quality of life and fatigue were not frequently reported. The RCTs ranged from 8 to 12 weeks' duration, so all data for benefit are based on only short-term studies. Follow-up effects were measured in only 14% of studies (N = 2). Whether the effects persist after the completion of exercise programmes is unknown.

Lastly, we were unable to assess the safety of exercise programmes, because adverse effects were not systematically monitored and reported in publications. Severe adverse events are rare with exercises but it can happen (for example fall). Exercise programmes are generally associated with minor adverse effects (muscle or joint pain, soreness) related to interventions (Kunutsor 2018). In our review, adverse effects were reported in one study, and only one event was associated with exercise programmes. Direct evidence for safety was not found, particularly for populations at risk, such as older people, or those with more severe AS. Data provided by Jacques 2014 suggested that mechanical strain can trigger inflammation, and cause bone degradation in mice. This result supports the need to systematically monitor adverse effects of exercise programmes, to determine if exercises are safe or if some adverse effects might be expected for example with a modification of exercise dosing / intensity (McGonagle 2014).

Quality of the evidence

We had concerns about risk of bias for all studies included: 79% (N = 11) had unclear allocation concealment; all failed to blind participants, staff, or outcome assessors; 43% had incomplete outcome data (N = 6), and 86% had unclear selective reporting, or a high risk of other bias (N = 12). Given the number of studies included in the review, we cannot rule out the existence of a small-study effect, explaining the magnitude of the positive results we found.

We considered statistically significant group differences between exercise programmes versus no intervention or usual care. Larger effects were found when exercise was compared to no intervention. For each comparison, the number of studies (< 10), and small samples (many studies were small, with < 100 participants) might have contributed to a low-power analysis. Low power is associated with bias (Button 2013). Most studies we included were assessed at high or unclear risk of bias, which suggests that the estimated effects might be over-estimated, and reduces the likelihood that they reflect a true effect. We cannot provide conclusions with a high level of confidence. The magnitude of the estimated effects may change with larger studies.

We only presented the findings of trials that reported the major outcomes of interest in Summary of findings for the main comparison and Summary of findings 2; and used the GRADE approach to assess the quality of the evidence examined for each outcome (Schünemann 2011b). Most of the evidence was downgraded to low or very low quality, based on three factors: risk of bias, inconsistency generated by heterogeneity, and imprecision with small trials and large confidence intervals.

Potential biases in the review process

We made all attempts to reduce the bias involved in the review process by including the best available evidence. All studies included were randomised trials. We conducted an extensive search of the literature in all relevant databases, but because two studies have not yet been incorporated, this may be a source of potential bias. Two review authors independently selected studies, extracted data, and assessed the risk of bias. For missing data, we attempted to extract data that were graphically displayed by using software tools (arohatgi.info/WebPlotDigitizer/index.html), or to systematically seek information from authors of the included studies.

The review itself has some limitations. We could not determine whether participants who received usual care also had exercises, because the included studies poorly described the content of usual care interventions. Participants in the usual care group could have practiced exercises, which could explain why a smaller effect size was always found when comparing exercise programmes to usual care. A possible explanation could also be the result of performance bias, due to lack of blinding.

We found wide variations among the trials, likely related to different exercise components. . Despite the pre-specification stated in the protocol, we could not perform subgroup analyses to explore heterogeneity for factors, such as supervision, modalities of exercises, or participant characteristics. Lastly, the number of included studies was too small.

Agreements and disagreements with other studies or reviews

Different systematic reviews have examined the effects of exercise programmes in people with AS. None included all of the RCTs we identified, all of which compared the effects of exercise programmes to no intervention or usual care.

Dagfinrud 2008 included 11 RCTs and quasi-randomised trials (763 participants). Only four studies compared exercise programmes with no intervention (Ince 2006; Kraag 1990; Lim 2005; Sweeney 2002). The other included studies compared different modalities of exercise programmes. The systematic review did not meta-analyse the results of the comparator groups. Only the effects of individual studies were reported. The authors reported low-quality evidence for effects on spinal mobility and physical function. An update was performed by Dagfinrud 2011, which included one additional study. The authors included the same four previously included

studies that compared different types of exercise programmes to no intervention. They did not meta-analyse the results.

Van den Berg 2012 performed a systematic review that included randomised and uncontrolled trials (cohort studies, case–control studies, and cross-sectional studies), which evaluated any type of non-pharmacological intervention. They concluded that exercise programmes were better than no intervention. This review included only one of the 14 trials included in our review (Widberg 2009).

O'Dwyer 2014 included randomised and quasi-randomised trials. This systematic review included only five of the 14 RCTs included in our review, and concluded that therapeutic exercise improved physical function, joint mobility, and cardiorespiratory function, and ameliorated patient-assessed disease activity, pain, and stiffness compared with controls. They assessed the evidence to be of moderate quality.

A recent systematic review compared specific pulmonary exercise programmes to conventional exercise or no intervention (Saracoglu 2017). The authors included eight RCTs or controlled trials. Two of the trials were included in our review, but the other six were excluded, because they did not meet our inclusion criteria (Altan 2012; Ince 2006). Evidence showed that exercise improved functional capacities and pulmonary functions, but the authors did not provide a critical appraisal of available studies. They did not conduct a meta-analysis.

Three systematic reviews examined the effect of exercises combined with stabilised TNF blocker therapy versus patients with AS stabilised by TNF blocker therapy (Giannotti 2014; Liang 2015; Lubrano 2015). These systematic reviews included non-randomised controlled trials, observational studies, and abstracts, which evaluated spa exercise therapy combined with stabilised TNF blocker therapy. The authors concluded that exercises combined with stabilised TNF blocker therapy might reduce patient-assessed disease activity, improve function and quality of life compared with biologic therapy alone. Our findings are consistent with previous reviews and guidelines that found short-term effects of exercise programmes (Regel 2017; Ward 2016).

AUTHORS' CONCLUSIONS

Implications for practice

We found moderate- to low-quality evidence indicating that exercise programmes compared to no intervention probably slightly improve function, may reduce pain (important clinical benefit), and probably slightly reduce patient-assessed disease activity, measured after the completion of the exercise programmes. Whether there was an effect on spinal mobility and fatigue is uncertain. We found moderate- to low-quality evidence that compared with usual care (including physiotherapy, medication, or self-management), exercise programmes probably have little or no effect on improving function and reducing pain, and may have little or no difference on patient-assessed disease activity. We are uncertain whether exercise programmes improve spinal mobility.

Readers should understand that we are uncertain of the potential for harm from exercise programmes, because of the limited number of studies reporting AEs, and the low rate of events.

We are unable to distinguish the best type of exercise, its components, or its mode of delivery.

Implications for research

The evidence for some of the major outcomes was low or very low quality, so new studies could change the estimate effects.

This review has raised new questions to answer:

- The long-term effects of exercise programmes for people with AS, and whether they are clinically relevant are unclear.
- New trials should provide an accurate description of the content, dose, application, and adherence to the exercise interventions. The most effective components (e.g. supervised or home delivery) are unknown, as is the most effective dose, including frequency, intensity, and duration.
- AEs were rarely measured and reported in RCT reports. Whether exercise programmes produce harmful effects is difficult to determine. AEs may be worthwhile to investigate in people with more advanced or severe stages of the disease. Studies should systematically investigate and report AEs.
- Further studies should investigate the effect of exercise therapy in the early stages of the disease (even in the pre-radiographic stages). Exercise programmes should be evaluated at different stages of the disease. This evaluation would be useful to ascertain whether the use of biologic agents and rehabilitation programmes in people with newly-diagnosed, or early AS, are effective to prevent deformity and disability.
- Studies of the effects of TNF blockers combined with exercise programmes, with cost-effectiveness, are needed.
- Cost-effectiveness of interventions should be evaluated.

Future studies should use the Consensus on Exercise Reporting Template, or the CONSORT Template for Intervention Description and Replication, to improve the description of exercise programmes and facilitate their application in clinical practice.

ACKNOWLEDGEMENTS

The authors thank L Trinquart for statistical expertise, the Cochrane Musculoskeletal Review Group editorial team for helpful support, and T Rader, D Wendling, and the reviewers for editorial comments. We also thank Laura Smales for her suggestions for translation and formatting.

REFERENCES

References to studies included in this review

Altan 2012 *{published data only}*

Altan L, Korkmaz N, Dizdar M, Yurtkuran M. Effect of Pilates training on people with ankylosing spondylitis. *Rheumatology International* 2012 Jul;**32**(7):2093-9. [PUBMED: PMID: 21499876]

Dönmez 2014 *{published data only}*

Dönmez U, Ozturk C, Kocanaogullari H, Gucenmez S, Hepguler S. Do physical therapy modalities have additional benefit over exercise therapy in the management of ankylosing spondylitis? A randomized controlled trial. *Annals of the Rheumatic Diseases* 2014;**73**(Suppl2):727. [SAT0369]

Gallinaro 2016 *{published data only}*

Gallinaro AL. Effects of Two Exercise Therapy Programs with and without Elastic Resistance in Ankylosing Spondylitis Patients [Doctoral Thesis]. São Paulo (Brazil): Faculdade de Medicina, Universidade de São Paulo, 2016. [USP/FM/DBD-199/16; www.teses.usp.br/teses/disponiveis/5/5169/tde-04112016-150051/fr.php]

NCT01690273. Exercise therapy program in ankylosing spondylitis patients. clinicaltrials.gov/ct2/show/NCT01690273 (first received 21 September 2012).

Garcia 2015 *{published data only}*

Fernández García R, Sánchez Sánchez L de C, López Rodríguez M del M, Sánchez Granados G. Effects of an exercise and relaxation aquatic program in patients with spondyloarthritis: a randomized trial. *Medicina Clínica (Barc)* 2015 Nov 6;**145**(9):380-4. [PUBMED: PMID: 25639496]

Ince 2006 *{published data only}*

Ince G, Sarpel T, Durgun B, Erdogan S. Effects of a multimodal exercise program for people with ankylosing spondylitis. *Physical Therapy* 2006;**Jul; 86**(7):924-35. Erratum in: Physical Therapy. 2006 Oct;86(10):1452. [PUBMED: PMID: 16813473]

Kjeken 2013 *{published data only}*

Kjeken I, Bø I, Rønningen A, Spada C, Mowinckel P, Hagen KB, et al. A three-week multidisciplinary in-patient rehabilitation programme had positive long-term effects in patients with ankylosing spondylitis: randomized controlled trial. *Journal of Rehabilitation Medicine* 2013;**Mar; 45**(3):260-7. [PUBMED: PMID: 23138412]

Kraag 1990 *{published data only}*

Kraag G, Stokes B, Groh J, Helewa A, Goldsmith C. The effects of comprehensive home physiotherapy and supervision on patients with ankylosing spondylitis-a randomized controlled trial. *The Journal of Rheumatology* 1990;**Feb; 17**(2):228-33. [PUBMED: PMID:2181127]

* Kraag G, Stokes B, Groh J, Helewa A, Goldsmith CH. The effects of comprehensive home physiotherapy and supervision on patients with ankylosing spondylitis – an 8-month follow-up. *The Journal of Rheumatology* 1994 Feb;**21**(2):261-3. [PUBMED: PMID:8182634]

Lim 2005 *{published data only}*

Lim HJ, Moon YI, Lee MS. Effects of home-based daily exercise therapy on joint mobility, daily activity, pain, and depression in patients with ankylosing spondylitis. *Rheumatology International* 2005;**Apr; 25**(3):225-9. [PUBMED: PMID: 15650833]

Masiero 2011 *{published data only}*

Masiero S, Bonaldo L, Pigatto M, Lo Nigro A, Ramonda R, Punzi L. Rehabilitation treatment in patients with ankylosing spondylitis stabilized with tumor necrosis factor inhibitor therapy: a randomized controlled trial. *The Journal of Rheumatology* 2011 Jul;**38**(7):1335-42. [PUBMED: PMID: 21459942]

Rodriguez-Lozano 2013 *{published data only}*

Rodríguez-Lozano C, Juanola X, Cruz-Martínez J, Peña-Arrébola A, Mulero J, Gratacós J, Collantes E, Spondyloarthropathies Study Group of the Spanish Society of Rheumatology. Outcome of an education and home-based exercise programme for patients with ankylosing spondylitis: a nationwide randomized study. *Clinical and Experimental Rheumatology* 2013 Sep-Oct;**31**(5):739-48. [PUBMED: PIMD: 23899791]

Souza 2017 *{published data only}*

Souza MC, Jennings F, Morimoto H, Natour J. Swiss ball exercises improve muscle strength and walking performance in ankylosing spondylitis: a randomized controlled trial. *Revista Brasileira de Reumatologia [English edition]* 2017 Jan - Feb;**57**(1):45-55. [PUBMED: PMID: 28137402]

Sveaas 2014 *{published data only}*

Sveaas SH, Berg IJ, Provan SA, Semb AG, Hagen KB, Vøllestad N, et al. Efficacy of high intensity exercise on disease activity and cardiovascular risk in active axial spondyloarthritis: a randomized controlled pilot study. *PLoS One* 2014;**Sep 30; 9**(9):e108688. [PUBMED: PMID: 25268365]

Sweeney 2002 *{published data only}*

Sweeney S, Taylor G, Calin A. The effect of a home based exercise intervention package on outcome in ankylosing spondylitis: a randomized controlled trial. *The Journal of Rheumatology* 2002;**Apr; 29**(4):763-6. [PUBMED: PMID: 11950019]

Widberg 2009 *{published data only}*

Widberg K, Karimi H, Hafström I. Self- and manual mobilization improves spine mobility in men with ankylosing spondylitis – a randomized study. *Clinical Rehabilitation* 2009;**Jul; 23**(7):599-608. [PUBMED: PMID: 19403551]

References to studies excluded from this review

Ciprian 2013 *{published data only}*

Ciprian L, Lo Nigro A, Rizzo M, Gava A, Ramonda R, Punzi L, et al. The effects of combined spa therapy and rehabilitation on patients with ankylosing spondylitis being treated with TNF

inhibitors. *Rheumatology International* 2013, Jan;**33**(1):241-5. [PUBMED: PMID: 21947374]

Colina 2009 *{published data only}*

Colina M, Ciancio G, Garavini R, Conti M, Trotta F, Govoni M. Combination treatment with etanercept and an intensive spa rehabilitation program in active ankylosing spondylitis. *International Journal of Immunopathology and Pharmacology* 2009 Oct-Dec;**22**(4):1125-9. [PUBMED: PMID: 20074478]

Durmus 2009 *{published data only}*

Durmus D, Alayli G, Cil E, Canturk F. Effects of a home-based exercise program on quality of life, fatigue, and depression in patients with ankylosing spondylitis. *Rheumatology International* 2009 Apr;**29**(6):673-7. [PUBMED: PMID: 18985351]

Gunay 2012 *{published data only}*

Gunay MO, Bal S, Bayram KB, Harman E, Dalgic EE, Kocyitit H, et al. The effect of breathing and posture exercise on the clinical, functional status and disease related quality of life in patients with ankylosing spondylitis. *Medicine Science* 2012;**1**(2):103-17. [10.5455/medscience.2012.01.103-117]

Karahan 2016 *{published data only}*

Karahan AY, Tok F, Yildirim P, Ordahan B, Turkoglu G, Sahin N. The effectiveness of exergames in patients with ankylosing spondylitis: a randomized controlled trial. *Advances in Clinical and Experimental Medicine* 2016;**25**(5):931-6. [DOI: 10.17219/acem/32590]

Lee 2008 *{published data only}*

Lee EN, Kim YH, Chung WT, Lee MS. Tai chi for disease activity and flexibility in patients with ankylosing spondylitis – a controlled clinical trial. *Evidence-based Complementary and Alternative Medicine: ECAM* 2008 Dec;**5**(4):457-62. [PUBMED: PMID: 18955296]

Masiero 2015 *{published data only}*

Masiero S. A controlled clinical trial on balance parameters improvement after physical rehabilitation in ankylosing spondylitis. *Archives of Physical Medicine and Rehabilitation* 2015;**10**:e91. [CN-01126811]

References to studies awaiting assessment

Mesquita 2014 *{published data only}*

Mesquita C, Lopes S, Moreira EM, Lopes P, Cameiro A, Silva D, et al. Effect of home exercise program in quality of life in patients with ankylosing spondylitis. *Annals of the Rheumatic Diseases* 2014;**73**(suppl 1):A41. [DOI: 10.1136/annrheumdis-2013-205124.93]

Sveeas 2018 *{published data only}*

Sveaas SH, Berg IJ, Fongen C, Provan SA, Dagfinrud H. High-intensity cardiorespiratory and strength exercises reduced emotional distress and fatigue in patients with axial spondyloarthritis: a randomized controlled pilot study.. *Scandinavian Journal of Rheumatology* 2018 Mar;**47**(2):117-21. [DOI: 10.1080/03009742.2017.1347276; PUBMED: PMID: 28891743]

References to ongoing studies

ChiCTR-TRC-14004650 *{published data only}*

ChiCTR-TRC-14004650. The effect of traditional exercise "Baduanjin" for physical functioning of ankylosing spondylitis: a randomized, controlled, prospective study. www.chictr.org.cn/hvshowproject.aspx?id=10477 (last refreshed 02 May 2015).

NCT02098694 *{published data only}*

NCT02098694. Physiotherapy-led outpatient clinic for patients with spondyloarthritis. clinicaltrials.gov/ct2/show/NCT02098694 (first posted 28 March 2014).

Additional references

Abenhaim 2000

Abenhaim L, Rossignol M, Valat JP, Nordin M, Avouac B, Blotman F, et al. The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. *Spine* 2000;**25**(4 Suppl):1S-33S. [MEDLINE: 10707404]

Arem 2015

Arem H, Moore SC, Patel A, Hartge P, Berrington de Gonzalez A, Visvanathan K, et al. Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. *JAMA Internal Medicine* 2015 Jun;**175**(6):959-67. [DOI: 10.1001/jamainternmed.2015.0533; PUBMED: PMID: 25844730]

Baraliakos 2005

Baraliakos X, Landewé R, Hermann KG, Listing J, Golder W, Brandt J, et al. Inflammation in ankylosing spondylitis: a systematic description of the extent and frequency of acute spinal changes using magnetic resonance imaging. *Annals of the Rheumatic Diseases* 2005;**64**(5):730-4. [PUBMED: PMC1755467]

Baraliakos 2012

Baraliakos X, van den Berg R, Braun J, van der Heijde D. Update of the literature review on treatment with biologics as a basis for the first update of the ASAS/EULAR management recommendations of ankylosing spondylitis. *Rheumatology (Oxford)* 2012;**51**(8):1378-87. [PUBMED: 22427410]

Boonen 2004

Boonen A, Van Tubergen A, Van Der Linden S, Mihai C, Ottawa Methods Group. Spondyloarthropathies. In: Tugwell P, Shea B, Boers M, Brooks P, Simon LS, Strand V, et al. editor(s). Evidence-Based Rheumatology. London: BMJ Books, 2004:393-429.

Boonen 2006

Boonen A, van der Linden SM. The burden of ankylosing spondylitis. *The Journal of Rheumatology* 2006;**78**(Suppl):4-11. [PUBMED: 17042055.]

Bouchard 2012

Bouchard C, Blair S, Haskell W. Physical Activity and Health. Champaign, IL: Human Kinetics, Inc, 2012.

Boutron 2007

Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing non pharmacological treatments. *PLoS Med* 2007 Feb;**4**(2):e61. [PUBMED: PMID: 17311468]

Braun 2003

Braun J, Pham T, Sieper J, Davis J, van der Linden S, Dougados M, et al. International ASAS consensus statement for the use of anti-tumour necrosis factor agents in patients with ankylosing spondylitis. *Annals of the Rheumatic Diseases* 2003;**62**(9):817-24.

Braun 2007

Braun J, Sieper J. Ankylosing spondylitis. *Lancet* 2007;**369**(9570):1379-90.

Braun 2010

Braun J, van den Berg R, Baraliakos X, Boehm H, Burgos-Vargas R, Collantes-Estevez E. 2010 update of the ASAS/EULAR recommendations for the management of ankylosing spondylitis. *Annals of the Rheumatic Diseases* 2011;**70**(6):896-904. [PUBMED: PMC3086052]

Buchbinder 2015

Buchbinder R, Golmohammadi K, Johnston RV, Owen RJ, Homik J, Jones A, et al. Percutaneous vertebroplasty for osteoporotic vertebral compression fracture. *Cochrane Database of Systematic Reviews* 2015, Issue 4. [CENTRAL: CD006349; DOI: 10.1002/14651858.CD006349.pub2]

Button 2013

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience* 2013 May;**14**(5):365-76. [PUBMED: PMID: 23571845]

Cates 2008 [Computer program]

Cates CD. Visual Rx version 3.0. London (UK): Chris Cates, (accessed 16 July 2014).

Cohen 2006

Cohen JD, Cunin P, Farrenq V, Oniankitan O, Carton L, Chevalier X, et al. Estimation of the Bath Ankylosing Spondylitis Disease ActivityIndex cutoff for perceived symptom relief in patients with spondyloarthropathies. *Journal of Rheumatology* 2006 Jan;**33**(1):79-81.

Dagfinrud 2005

Dagfinrud H, Kjeken I, Mowinckel P, Hagen KB, Kvien TK. Impact of functional impairment in ankylosing spondylitis: impairment, activity limitation, and participation restrictions. *Journal of Rheumatology* 2005;**32**(3):516-23.

Dagfinrud 2008

Dagfinrud H, Kvien TK, Hagen KB. Physiotherapy interventions for ankylosing spondylitis. *Cochrane Database of Systematic Reviews* 2008, Issue 1. [DOI: 10.1002/14651858.CD002822.pub3]

Dagfinrud 2011

Dagfinrud H, Halvorsen S, Vøllestad NK, Niedermann K, Kvien TK, Hagen KB. Exercise programs in trials for patients with ankylosing spondylitis: do they really have the potential for effectiveness?. *Arthritis Care & Research* 2011 Apr;**63**(4):597-603. [PUBMED: PMID: 21452270]

Dean 2014

Dean LE, Jones GT, MacDonald AG, Downham C, Sturrock RD, Macfarlane GJ. Global prevalence of ankylosing spondylitis. *Rheumatology (Oxford)* 2014;**53**(4):650-7. [PUBMED: 24324212]

Deeks 2011

Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Deeks 2019

Deeks JJ, Higgins JPT, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. Cochrane Handbook for Systematic Reviews of Interventions. Version 6. Draft version (29 january 2019). Available from handbook.cochrane.org.

Dougados 2011

Dougados M, Baeten D. Spondyloarthritis. *Lancet* 2011;**377**(9783):2127-37.

Dubey 2008

Dubey SG, Leeder J, Gaffney K. Physical therapy in anti-TNF treated patients with ankylosing spondylitis. *Rheumatology (Oxford)* 2008 Jul;**47**(7):1100-1. [PUBMED: PMID: 18495822]

Fernández-de-Las-Peñas 2005

Fernández-de-Las-Peñas C, Alonso-Blanco C, Morales-Cabezas M, Miangolarra-PageJC. Two exercise interventions for the management of patients with ankylosing spondylitis: a randomized controlled trial. *American Journal of Physical Medicine & Rehabilitation* 2005 Jun;**84**(6):407-19. [PUBMED: 15905654]

Giannotti 2014

Giannotti E, Trainito S, Arioli G, Rucco V, Masiero S. Effects of physical therapy for the management of patients with ankylosing spondylitis in the biological era. *Clinical Rheumatology* 2014 Sep;**33**(9):1217-30. [PMID: 24797772]

GRADEpro GDT [Computer program]

McMaster University (developed by Evidence Prime). GRADEpro GDT. Version accessed. Hamilton (ON): McMaster University (developed by Evidence Prime), 2015.

Hagen 2012

Hagen KB, Dagfinrud H, Moe RH, Østerås N, Kjeken I, Grotle M, et al. Exercise therapy for bone and muscle health: an overview of systematic reviews. *BMC Medicine* 2012 Dec;**19**(10):167. [PUBMED: PMC3568719]

Haroon 2013

Haroon N, Inman RD, Learch TJ, Weisman MH, Lee M, Rahbar MH, et al. The impact of tumor necrosis factor α inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis and Rheumatism* 2013;**65**(10):2645-54. [PUBMED: 23818109]

Hayden 2012

Hayden JA, Cartwright J, van Tulder MW, Malmivaara A. Exercise therapy for chronic low back pain. *Cochrane Database of Systematic Reviews* 2012, Issue 4. [DOI: 10.1002/14651858.CD009790]

Higgins 2011a

Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2011b

Higgins JPT, Deek JJ, Altman DG. Chapter 16: Special topics in statistics. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2011c

Higgins JPT, Deeks JJ. editor(s). Chapter 7: Selecting studies and collecting data. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org. The Cochrane Collaboration.

Hoffmann 2014

Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE, Dixon-Woods M, McCulloch P, WyattJC, Chan AW, Michie S. Better reporting of interventions: template forintervention description and replication (TIDieR) checklist and guide. *BMJ* 2014 Mar 7;**348**:g1687. [10. 1136/bmj. g1687]

Hozo 2005

Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology* 2005 Apr;**20**(5):13. [PUBMED: PMID: 15840177]

Ioannidis 2007a

Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* 2007;**336**(7626):914-6.

Jacques 2014

Jacques P, Lambrecht S, Verheugen E, Pauwels E, Kollias G, Armaka M, et al. Proof of concept: enthesitis and new bone formation in spondyloarthritis are driven by mechanical strain and stromal cells. *Annals of the Rheumatic Diseases* 2014 Feb;**73**(2):437-45. [PUBMED: PMID: 23921997]

Kujala 2009

Kujala UM. Evidence on the effects of exercise therapy in the treatment of chronic disease. *British Journal of Sports Medicine* 2009 Aug;**43**(8):550-5. [PUBMED: 19406731]

Kunutsor 2018

Kunutsor S, Leyland S, Skelton D, James L, Cox M, Gibbons N, Whitney J, Clark E. Adverse events and safety issues associated with physical activity and exercise for adults with osteoporosis and osteopenia: A systematic review of observational studies and an updated review of interventional studies. *Journal of Fraitly, Sarcopenia & Frails* 2018 December;**3**(4):155-178.

Lefebvre 2011

Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Liang 2015

Liang H, Li WR, Zhang H, Tian X, Wei W, Wang CM. Concurrent intervention with exercises and stabilized tumor necrosis factor inhibitor therapy reduced the disease activity in patients with ankylosing spondylitis: a meta-analysis. *Medicine (Baltimore)* 2015 Dec;**94**(50):e2254. [PUBMED: PMC5058915]

Lubrano 2015

Lubrano E, Spadaro A, Amato G, Benucci M, Cavazzana I, Chimenti MS, et al. Tumour necrosis factor alpha inhibitor therapy and rehabilitation for the treatment of ankylosing spondylitis: a systematic review. *Seminars in Arthritis and Rheumatism* 2015 Apr;**44**(5):542-50. [PUBMED: PMID: 25450240]

Maxwell 2015

Maxwell LJ, Zochling J, Boonen A, Singh JA, Veras MM, Tanjong Ghogomu E, et al. TNF-alpha inhibitors for ankylosing spondylitis. *Cochrane Database of Systematic Reviews* 2015, Issue 4. [DOI: 10.1002/14651858.CD005468.pub2; PUBMED: PMID: 25887212]

McGonagle 2014

McGonagle D, Thomas RC, Schett G. Spondyloarthritis: may the force be with you?. *Annals of the Rheumatic Diseases* 2014 Feb;**73**(2):321-3. [PUBMED: PMID: 24220157]

Niedermann 2013

Niedermann K, Sidelnikov E, Muggli C, Dagfinrud H, Hermann M, Tamborrini G, et al. Effect of cardiovascular training on fitness and perceived disease activity in people with ankylosing spondylitis. *Arthritis Care & Research* 2013 Nov;**65**(11):1844-52. [PUBMED: 23836515]

O'Dwyer 2014

O'Dwyer T, O'Shea F, Wilson F. Exercise therapy for spondyloarthritis: a systematic review. *Rheumatology International* 2014 Jul;**34**(7):887-902. [PMID: 24549404]

Palla 2012

Palla I, Trieste L, Tani C, Talarico R, Cortesi PA, Mosca M, et al. A systematic literature review of the economic impact of

ankylosing spondylitis. *Clinical and Experimental Rheumatology* 2012;**30**(4 Suppl 73):S136-41. [PUBMED: 23072824]

Poddubnyy 2012

Poddubnyy D, Rudwaleit M, Haibel H, Listing J, Märker-Hermann E, Zeidler H, et al. Effect of non-steroidal anti-inflammatory drugs on radiographic spinal progression in patients with axial spondyloarthritis: results from the German Spondyloarthritis Inception Cohort. *Annals of the Rheumatic Diseases* 2012;**71**(10):1616-22. [PUBMED: 22459541]

Pradeep 2008

Pradeep DJ, Keat A, Gaffney K. Predicting outcome in ankylosing spondylitis. *Rheumatology (Oxford)* 2008;**47**(7):942-5. [PUBMED: 18492709]

Ramiro 2014

Ramiro S, van der Heijde D, van Tubergen A, Stolwijk C, Dougados M, van denBosch F, et al. Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. *Annals of the Rheumatic Diseases* 2014 Aug;**73**(8):1455-61. [PUBMED: PMID: 24812292]

Regel 2017

Regel A, Sepriano A, Baraliakos X, van der Heijde D, Braun J, Landewé R, et al. Efficacy and safety of non-pharmacological and non-biological pharmacological treatment: a systematic literature review informing the 2016 update of the ASAS/EULAR recommendations for the management of axial spondyloarthritis. *Rheumatic and Musculoskeletal Diseases Open* 2017;**3**(1):e000397.

Reveille 2012

Reveille JD, Ximenes A, Ward MM. Economic considerations of the treatment of ankylosing spondylitis. *The American Journal of the Medical Sciences* 2012;**343**(5):371-4. [PUBMED: PMC3613240.]

RevMan 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Rucker 2008

Rucker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Medical Research Methodology* 2008;**8**:79. [DOI: 10.1186/1471-2288-8-79]

Rudwaleit 2009

Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Annals of the Rheumatic Diseases* 2009 Jun;**68**(9):777-83. [PUBMED: PMID: 19297344]

Saracoglu 2017

Saracoglu I, Kurt G, Okur EO, Afsar E, Seyyar GK, Calik BB, et al. The effectiveness of specific exercise types on cardiopulmonary functions in patients with ankylosing spondylitis: a systematic review. *Rheumatology International* 2017 Mar;**37**(3):409-21. [PUBMED: PMID: 27837263]

Saunders 2016

Saunders DH, Sanderson M, Hayes S, Kilrane M, Greig CA, Brazzelli M, et al. Physical fitness training for stroke patients. *Cochrane Database of Systematic Reviews* 2016, Issue 3. [DOI: 10.1002/14651858.CD003316.pub6; PUBMED: 27010219]

Schünemann 2011a

Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: Presenting results and 'Summary of findings' tables. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Schünemann 2011b

Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Sieper 2009

Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. *Annals of the Rheumatic Diseases* 2009;**68**(Suppl 2):ii1-44.

Singh 2013

Singh B, Upadhyay A, Garg M, Mahajan A. Effects of physical and rehabilitative treatment on quality of life of ankylosing spondylitis patients. *The Internet Journal of Orthopedic Surgery* 2013;**21**(2):1-5.

Slade 2016

Slade SC, Dionne CE, Underwood M, Buchbinder R, Beck B, Bennell K, et al. Consensus on Exercise Reporting Template (CERT): modified Delphi study. *Physical Therapy* 2016 [Epub 2016 May 5];**96**(10):1514-24. [DOI: 10.2522/ptj.20150668; PUBMED: 27149962]

Slobodin 2015

Slobodin G, Eshed I. Non-radiographic axial spondyloarthritis. *Israel Medical Association Journal* 2015;**17**(12):770-6. [PUBMED: PMID: 26897981]

Tubach 2012

Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. *Arthritis Care & Research* 2012 Nov;**64**(11):1699-707. [PUBMED: 22674853]

Van den Berg 2012

Van den Berg R, Baraliakos X, Braun J, van der Heijde D. First update of the current evidence for the management of ankylosing spondylitis with non-pharmacological treatment and non-biologic drugs: a systematic literature review for the ASAS/EULAR management recommendations in ankylosing spondylitis. *Rheumatology (Oxford)* 2012;**51**(8):1388-96.

Van den Berg 2013

Van den Berg R, de Hooge M, Rudwaleit M, Sieper J, van Gaalen F, Reijnierse M, et al. ASAS modification of the Berlin algorithm for diagnosing axial spondyloarthritis: results from the SPondyloArthritis Caught Early (SPACE)-cohort and from the Assessment of SpondyloArthritis international Society (ASAS)-cohort. *Annals of the Rheumatic Diseases* 2013;**72**(10):1646-53. [PUBMED: PMID: 23139266]

Van der Heijde 1997

Van der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. *Journal of Rheumatology* 1997;**24**(11):2225-9. [PUBMED: 9375888]

Van der Heijde 1999

Van der Heijde D, van der Linden S, Bellamy N, Calin A, Dougados M, Khan MA. Which domains should be included in a core set for endpoints in ankylosing spondylitis? Introduction to the ankylosing spondylitis module of OMERACT IV. *Journal of Rheumatology* 1999;**26**(4):945-7. [PUBMED: 10229424]

van der Heijde 2008

Van der Heijde D, Landewé R, Baraliakos X, Houben H, van Tubergen A, Williamson P, et al. Radiographic findings following two years of infliximab therapy in patients with ankylosing spondylitis. *Arthritis and Rheumatism* 2008;**58**(10):3063-70.

Van der Linden 1984

Van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis and Rheumatism* 1984;**27**(4):361-8.

Vidoni 2015

Vidoni ED, Johnson DK, Morris JK, Van Sciver A, Greer CS, Billinger SA, et al. Dose-response of aerobic exercise on cognition: a community-based, pilot randomized controlled trial. *PLoS One* 2015 Jul 9;**10**(7):e0131647. [DOI: 10.1371/journal.pone.0131647; PUBMED: PMID: 26158265]

Vliet 2009

Vliet Vlieland TP, Li LC. Rehabilitation in rheumatoid arthritis and ankylosing spondylitis: differences and similarities. *Clinical and Experimental Rheumatology* 2009;**27**(4 Suppl 55):S171-8.

Wang 2009

Wang CY, Chiang PY, Lee HS, Wei JC. The effectiveness of exercise therapy for ankylosing spondylitis: a review. *International Journal of Rheumatic Diseases* 2009;**12**(3):207-10.

Ward 2016

Ward MM, Deodhar A, Akl EA, Lui A, Ermann J, Gensler LS, et al. American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network 2015 Recommendations for the treatment of ankylosing spondylitis and non-radiographic axial spondyloarthritis. *Arthritis Care & Research* 2016 Feb;**68**(2):151-66. [PUBMED: PMID: 26401907]

Webers 2016

Webers C, Essers I, Ramiro S, Stolwijk C, Landewé R, van der Heijde D, et al. Gender-attributable differences in outcome of ankylosing spondylitis: long-term results from the Outcome in Ankylosing Spondylitis International Study. *Rheumatology (Oxford)* 2016 Mar;**55**(3):419-28. [DOI: 10.1093/rheumatology/kev340; PUBMED: PMID: 26385369]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Altan 2012

Methods	RCT with two groups No intention-to-treat analysis
Participants	Location: Turkey Randomised: 55 Analysed: 53 Age: 45 years Gender: 55% men

Altan 2012 *(Continued)*

	Recruited: from a rheumatology clinic, mono-centre Inclusion criteria: AS according to the modified New York criteria, regular follow-up protocol in the clinic Exclusion criteria: systemic problem contraindicating exercising, peripheral arthritis, total spinal ankylosis, ESR over 50 mm/h, CRP more than 10 times the normal value, changed treatment regimens during the 2 months prior to the study Severity and duration of the disease: the duration of disease was 2 to 22 years (mean: 9) Coexisting medication treatment: 31% received NSAID, 32% sulphasalazine, and 21% biological agent treatment regularly, while 17% did not use regular medication. No change of medication during the study.
Interventions	Exercise group (N = 29) • Monomodal programme with exercises • Exercise components: Pilate exercise programme comprising 9 modules: postural education, search for neutral position, sitting exercise, pain relief exercises, stretching exercises, proprioceptive improvement exercises, and breathing education • Dose: 60min * 3times a week * 12weeks duration • Equipment: resistive bands and Pilate balls • Delivery mode: Supervised, unclear if the delivery was in group or individually. Propably in a rheumatology clinic but it was not clearly reported • Provider: by certified trainers • Tailoring: "difficulty levels of the exercises were adapted to the physical capacity of the patients" Control group (N = 24) • Usual care 'previous standard treatment' • Participants may practiced physical activity Adherence • Compliance: high; one person in each group quit the study • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0, final point = 12, follow-up = 24 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Spinal mobility: BASMI (VAS 10-cm scale), Chest expansion (cm) • Safety: not reported as an outcome Minor outcomes • Quality of life: ASQoL (0 to 18)
Notes	Number of missing participants: 2 (4%) Dropouts or withdrawals: n = 2; 1 in the exercise group and 1 in the control group Adverse effects: n = 1 (2%) in exercise group complained of increased back pain Sample size calculation: not reported Funding source: not available Declaration of interest: authors declared no conflict of interest

Altan 2012 *(Continued)*

	Protocol registration number not found
	Additional information provided after e-mail contact

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Comment: adequate method of random sequence generation Quote: "The participants were assigned randomly into two groups using random number table by the researcher other than the one who performed the evaluation throughout the study."
Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the intervention allocated Quote: "The participants were fully informed about the nature and purpose of the study."
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded Quote: "the same researcher who was totally unaware of the groups the participants belonged to, and all participants were requested not to give information to the examiner about their treatment protocol."
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: as treated analysis was carried out. Few balanced dropouts (1 in each group). May have a minor influence on the estimate of effect size.
Selective reporting (reporting bias)	Unclear risk	Comment: no registration protocol number reported. Outcomes listed in the method section were all reported in the results section. Safety was not reported as an outcome. Major and minor outcomes were specified.
Other bias	Unclear risk	Comment: attendance and compliance were not reported. Small sample size. Baseline performances were similar on major outcomes. Rates of medications did not differ between the two groups. No power sample calculation. Funding and conflict of interest were not declared.

Dönmez 2014

Methods	RCT with three groups
	intention-to-treat analysis performed
Participants	Location: Turkey
	Randomised: 77
	Analysed: 77
	Age: 42 years
	Gender: 53% men

Dönmez 2014 *(Continued)*

	Recruited: unclear Inclusion criteria: having a diagnosis of AS (meeting the Modified New York, ASAS criteria, or both, for axial SpA). Not doing any regular physical exercise in the last 6 months Exclusion criteria: severe cardiac, neurologic, or cognitive dysfunction, history of surgical intervention for spinal cord, pregnancy, severe psychiatric disorder, fractures due to secondary osteoporosis, being wheelchair-bound Severity and duration of the disease: not reported Coexisting medication treatment: no interruption or change of medication during the study
Interventions	Exercise group (N = 25) • Monomodal programme with exercises • Exercise components: global postural re-education exercises including breathing, stretching, and strength; limited description of the exercise programme components; warm-up for 10 minutes; cooling off for 10 minutes • Dose: 50 min/session, 5 days a week, for 3 weeks • Equipment: no description • Delivery mode: group session supervised; setting not reported • Provider: by an experienced physiotherapist. • Tailoring: no description Control group (N = 26) • No intervention group • control group which received only medical therapy • They were taught to continue their usual treatment. Adherence • Compliance: no information reported • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0, final point = 3, follow-up = 12 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Spinal mobility: BASMI (VAS 10-cm scale), • Pain intensity (VAS) • Safety: not reported as an outcome Minor outcomes • Quality of life: SF-36
Notes	Unclear if participants in the control group could practice community-based exercises Number of missing participant: n = 0; unclear reporting Dropouts or withdrawals: n = 0; unclear reporting Adverse events: not reported Adverse effects: not reported Sample size calculation: not reported

Dönmez 2014 *(Continued)*

	Funding source: not available
	Declaration of interest: authors declared no conflict of interest
	Protocol registration number not found
	Additional information provided after e-mail contact

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "patients [.] were allocated to three groups using a random numbers table"
Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the intervention allocated
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: the study did not provide information about missing data. After e-mail contact, the authors stated that no data were missing
Selective reporting (reporting bias)	Unclear risk	Comment: no registration protocol number reported; no information provided in the publication. After e-mail contact, the authors stated that all the outcomes were reported.
Other bias	Unclear risk	Comment: attendance and compliance were not reported. Inbalance at baseline was reported. No sample size calculation. Limited information in the published abstract. Funding and conflict of interest were not stated.

Gallinaro 2016

Methods	RCT with three groups
	Intention-to-treat: unclear
Participants	Location: Brazil
	Randomised: 55
	Analysed: 55
	Age: 48.7 years
	Gender: 87% men
	Recruited: from outpatient centre of São Paulo University, clinics of São Paulo.
	Inclusion criteria: having a diagnosis of AS (meeting the Modified New York); BASDAI < 4; stable medication for 6 months, not doing any regular physical exercise in the last 6 months. Functional class I to III.

Gallinaro 2016 *(Continued)*

	Exclusion criteria: change of medication during the study, doing exercise during the study period, important functional or walking limitation, cardiac problem contraindicating exercising, fibromyalgia, pain VAS over 8 Severity and duration of the disease: the mean BASDAI was 2.2; disease duration was 18.2 years Coexisting medication treatment: NSAIDs, biological treatment, DMARDs
Interventions	Exercise group (N = 37) • Monomodal programme with exercises • Exercise components: mobility exercise without (exercise group 1) or with resistance (exercise group 2) • 30 exercises including stretching and active joint motion • Dose: 30 min, 2 days a week, for 16 weeks • Equipment: no description • Delivery mode: group session supervised; setting not reported • Provider: by a physiotherapist • Tailoring: Control group (N = 18) • No intervention group • The control group received only medical therapy • No exercise practiced during 4 months Adherence • Compliance: not reported • Attendance: monitored; mean attendance to the 32 planned sessions was 49.3% (27.4
Outcomes	Time points (weeks): baseline = 0 , final point = 16 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Spinal mobility: BASMI (VAS 10-cm scale), • Schober test (tape in cm) • Fingertip to floor (tape in cm) • Chest expansion (tape in cm) • Pain intensity (VAS) • Stifness Safety: reported as an outcome Minor outcomes • Quality of life: SF-12 • Maastricht Ankylosing Spondylitis Enthesitis Score (MASES)
Notes	Number of missing participant: n = 0 Dropouts or withdrawals: n = 0 Adverse events: reported Averse effects: reported Sample size calculation: not reported

Gallinaro 2016 *(Continued)*

	Funding source: not available Declaration of interest: not available Protocol registration number: NCT01690273 We did not contact the author; additional information was found in a thesis manuscript (www.teses.usp.br/teses/disponiveis/5/5169/tde-04112016-150051/fr.php)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Comment: not sufficient information provided Quote: "The AS patients were randomly assigned into three groups".
Allocation concealment (selection bias)	Unclear risk	Comment: incomplete description found about allocation concealment Quote "em 3 grupos de pesquisa atraves de envelope pardo lacrado".
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the intervention allocated
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: all patients were included for analysis; Tables 2 & 3
Selective reporting (reporting bias)	Low risk	Comment: The study protocol and the results were posted on ClinicalTrials.gov
Other bias	Unclear risk	Comment: small number of subjects; no power sample size calculation; conflicts of interest were not declared

Garcia 2015

Methods	RCT with two groups intention-to-treat analysis performed
Participants	Location: Spain Randomised: 30 Analysed: 30 Age: 47 years Gender: 53% men Recruited: from association of arthritis and spondylitis Inclusion criteria: having a diagnosis of AS by a rheumatologist (meeting the modified New York, ASAS criteria, or both, for axial SpA), take only NSAIDS a month before, and during the study period

Garcia 2015 *(Continued)*

	Exclusion criteria: cardiovascular disease Severity and duration of the disease: mean time since diagnosis = 7 years Coexisting medication treatment: NSAIDS (100%)
Interventions	Exercise group (N = 15) • Monomodal programme with aquatic exercises • The programme included relaxation technique (10min), breathing technique (10min), active joint exercises (5min and 5s duration each movement), strength-resistance exercise with hip muscles at 50 to 70% of maximal strength during 15min, ending with aerobic exercise (20min) • Dose: 60 min, 3 days a week, 8-week duration • Equipment: aquatic; no description • Delivery mode: supervised, but unclear if it was in group or individualised sessions; setting: at the university sport facility • Provider: not described • Tailoring: strength was adjusted on the maximal strength and on the maximal heart rate Control group (N = 15) • no intervention Adherence • Compliance: no information reported • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0 , final point = 8 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Fatigue: BASDAI (VAS 10-cm scale) • Safety: not reported as an outcome Minor outcomes • Quality of life: SF-12
Notes	Number of missing participant: n = 0 Dropouts or withdrawals: n = 0 Adverse events: not reported Averse effects: not reported Sample size calculation: not reported Funding source: not available Protocol registration number not found Declaration of interest: authors declared no conflict of interest

Risk of bias

Bias	Authors' judgement	Support for judgement

Garcia 2015 *(Continued)*

Random sequence generation (selection bias)	Low risk	Quote: "individuals were randomised using random numbers"
Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the intervention allocated
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: all patients were included in the analysis
Selective reporting (reporting bias)	Unclear risk	Comment: no registration protocol number reported. Outcomes listed in the methods section were all reported in the results section. Safety was not reported as an outcome.
Other bias	Unclear risk	Comment: attendance and compliance were not reported; imbalance at baseline was reported; no sample size calculation

Ince 2006

Methods	RCT with two groups intention-to-treat analysis unclear
Participants	Location: Turkey Randomised: 30 Analysed: unclear Age: 35 years Gender: 60% men Setting: referred by their physician to the university department hospital, monocentre Inclusion criteria: AS according to the modified New York criteria Exclusion criteria: not reported Severity and duration of the disease: 4 participants with a stage II, 26 with a stage I (according to the modified New York criteria), mean disease duration = 9 years Coexisting medication treatment: NSAIDS + Sulfasalazine (100%)
Interventions	Exercise group (N = 15) • Monomodal programme with exercises • Exercise components: aerobic, stretching, and pulmonary exercises • Dose: 50 min, 3 times/weeks,12-wk duration • Aerobic exercises: low-intensity training based on MaxHR frequency, based on metronome

Ince 2006 *(Continued)*

- Stretching exercises: 14 exercises during the warm up and the cool down
- Pulmonary exercises: twice the normal rate of inspiration
- Warm-up: 10 minutes of step exercises (each motion repeated 10 times) 5 minutes of stretching exercises
- Main period: 20 minutes of step exercises (each motion repeated 10 times).
- Cool-down: 10 min of pulmonary exercises 5 min of stretching exercises.

- Equipement: no description
- Delivery mode: with supervision; unclear if in group or individual; setting: at the department of Physical therapy, Cukurova Hospital
- Provider: by "a doctorally trained exercise instructor with 10 years of experience"
- Tailoring: only intensity of aerobic exercise was adjusted for individual physical capacity, calculated with the Karvoven formula.

Control group (N = 15)

- No exercise group
- Participants received only their medical treatment; they did not practice any exercise

Adherence

- Compliance: not sufficient information; "all the subjects regularly attended the exercise programme"
- Attendance: no information reported

Outcomes	Time points (weeks): baseline = 0, final point = 12 Major outcomes • Spinal mobility: Schober Flexion Test (in cm) 10 cm above, occiput-to-wall distance (in cm), chest expansion (in cm) • Safety: not reported as an outcome Minor outcomes
Notes	Number of missing participant: not reported Dropouts or withdrawals: not reported Adverse events: not reported Adverse effects: not reported sample size calculation: not provided Funding source: this study was supported by the Research Project Unit of Cukurova University, Adana, Turkey (Project No: SBE2002D12) Protocol registration number not found Declaration of interest: not reported Additional informations provided after e-mail contact

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Comment: more information provided after e-mail contact: "randomisation with coin tossing" Quote: "the 30 subjects with AS were randomly divided into two groups"

Ince 2006 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: inadequate blinding; personnel unblinded Quote: "The exercise instructor was blinded to physiologic measures."
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: the study did not provide information about missing data; additional information provided after e-mail contact. The authors stated that all the patients completed the study and were analysed.
Selective reporting (reporting bias)	Unclear risk	Comment: no registration protocol number reported. Outcomes listed in the methods section were all reported in the results section; no major and minor outcomes designation
Other bias	Unclear risk	Comment: small number of subjects; main and recommended outcomes not used (no BASDAI, BASMI, or BASDAI), no power sample size calculation. Conflicts of interest were not declared.

Kjeken 2013

Methods	RCT with two groups No Intention-to-treat analysis.
Participants	Location: Norway Randomised: 100 Analysed: 95 Age: 49 years Gender: 67% men Recruited: from the hospital outpatient clinic and rheumatology department, multicentre (two different hospitals) Inclusion criteria: age between 18 and 65 years, AS according to the modified New York criteria, regular follow-up protocol in the clinic, BASDAI ≥ 40 mm Exclusion: coronary heart disease; pregnancy; impaired function due to other significant medical problems; surgery or rehabilitation within the last 6 months; or cognitive or mental impairment. "In addition, participants in the control group were excluded at the 4-month control if they reported participation in multidisciplinary rehabilitation after baseline assessment. Also, participants in both groups were excluded at the 12-month control if they had started biological therapy during the trial period, or reported multidisciplinary rehabilitation after the 4-month assessment." Severity and duration of the disease: disease duration = 15.5 yrs Coexisting medication treatment: analgesics (30.5%), NSAIDs (75.8%), biological therapy (7.4%), DMARDS (4.2%)

Kjeken 2013 *(Continued)*

Interventions	Exercise group (N = 46) • Multidsciplinary programme with self management and exercise • exercise components: strength, mobility, cardio and respiratory fitness. In addition, participants received individual physiotherapy when needed, including manual techniques. • Dose: 132 minutes, 3 sessions a week, for 3 weeks • Equipment: in the gym, in a hot water pool, and outdoor physical activities; no description • Delivery mode: no description; setting: no information • Provider: not reported • Tailoring: an individualised plan for the rehabilitation stay was developed, including patient-specific long- and short-term goals Control group (N = 49) • Usual care which could include community based physiotherapy, self-management, or both, in terms of physical activity and exercises". Adherence • Compliance: no information reported • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0 , final point = 16 , follow-up = 48 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Spinal mobility: BASMI (VAS 10-cm scale) • Safety: not reported as an outcome Minor outcomes • Quality of life: SF-363
Notes	Unclear if participants in the control group could practice community-based exercises Number of missing participants: 10 participants were excluded prior to follow-up assessments, most frequently due to participation in rehabilitation programmes at other centres in the trial period (n = 6 in the control group, and 4 in the exercise group). A total of 80% and 63% completed 4- and 12-month assessments in the rehabilitation group, vs 71% and 61% in the control group Dropouts or withdrawals: n = 25 (n = 12 in the exercise group; n = 13 in the control group) Adverse events: n =3 (3%) Averse effects: not reported Sample size calculation: reported; needed 50 patients in each group Funding source: this work was supported by Health South-East, Norway, grant number 2006077 Protocol registration number: ISRCTN75685576 Declaration of interest: not reported

Risk of bias

Bias	Authors' judgement	Support for judgement

Kjeken 2013 *(Continued)*

Random sequence generation (selection bias)	Low risk	Comment: a computer-based sequence generation was used Quote: "Patients were randomly assigned to the intervention or control group. A statistician not involved in the study made a computer-generated randomisation list."
Allocation concealment (selection bias)	Low risk	Comment: additional information after e-mail contact. The authors reported that envelopes were numbered Quote: "Concealed, opaque envelopes, prepared by a secretary, were used to allocate the patients [...] open after baseline assessments".
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: personnel unblinded Quote: "In this trial, the patients and therapists delivering the intervention were aware of the treatment assigned."
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: high rate of drop-out (28%); no intention-to-treat analysis; no imputation techniques used
Selective reporting (reporting bias)	High risk	Comment: differences found between outcomes stated in the protocol and outcomes reported in the publication Quote: "Biological signs of inflammation: erythrocyte sedimentation rate and C-reactive protein" were not reported. We could not incorporate data from the SF-36 scale in the meta-analysis
Other bias	Unclear risk	Comment: attendance and compliance not reported. No systematic reporting of adverse events. Outcomes reported a medium- and long-term follow-up. Power sample size calculation. Data adjusted for baseline values. Conflicts of interest were not declared

Kraag 1990

Methods	RCT with two groups modified intention-to-treat analysis
Participants	Location: Canada Randomised: 53 Analysed: 52 Age: 38 years (range: 19 to 73) Gender: 79% men Recruited: from the Arthritis Society home physiotherapy service Inclusion criteria: AS according to the New York criteria, English comprehension, absence of corticotherapy for at least 3 months, absence of immunotherapy for at least 6 months pre study, stable clini-

Kraag 1990 *(Continued)*

	cal status and drug therapy, ARA functional class 1, 2, or 3, no surgery anticipated in the next 4 months, if female, practising reliable contraception and not pregnant Exclusion criteria: more than 10% loss of flexion in either hip joint Severity and duration of the disease: no information mean pain at baseline (in mm): 35 morning stiffness: 75% Coexisting medication treatment: no change of medication treatment during the study
Interventions	Exercise group (N = 25) • Multi disciplinary programme with exercise and education • Exercise component: cold, heat, posture, flexibility, strength • Dose: between 8 to 16 sessions during the 4 months, limited information • Equipement: no information • Delivery mode: home physiotherapy in one-to-one education strategy • Provider: by an Arthritis Society physiotherapist • Tailoring: not reported Control group (N = 27) • No exercise group • "Participants did not received physiotherapy and education during the 4-month study period". Adherence • Compliance: no information reported • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0 , final point = 16 Major outcomes • Pain: VAS (0 to 100) • Spinal mobility: Schober test (cm), occiput-to-wall distance (cm) • Safety: not reported as an outcome Minor outcomes
Notes	Number of missing participant: 5 (9%) dropouts or withdrawals: n = 5 (exercise group = 4, control group = 1) adverse events: • 3 experienced a disease flare requiring medical intervention • 1 was in a leg cast • 1 medication change due to drug side effect adverse effect: not reported sample size calculation: not reported Funding source: supported by Nathonal Health Research and development programme, Health and Welfare Canada FGrant number 6606-2385-43 Protocol registration number not found

Kraag 1990 *(Continued)*

	Declaration of interest: not reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Comment: sufficient information not provided Quote: "fifty-three patients with ankylosing spondylitis (AS) were randomly allocated".
Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: modified intention-to-treat analysis. Unbalanced rate of dropouts (exercise group 15%; control group 4%) as 1 subject was missing for the final analysis. Reasons for dropouts were unclear.
Selective reporting (reporting bias)	Unclear risk	Comment: no registration protocol number reported. Major and minor outcomes were specified. Outcomes listed in the method section were all reported in the results section.
Other bias	Unclear risk	Comment: small number of participants, no information on baseline disease severity between the 2 groups, adherence and compliance to treatments was not monitored. No power sample size calculation. Conflicts of interest were not declared

Lim 2005

Methods	RCT with two groups No intention-to-treat analysis
Participants	Location: Korea Randomised: 58 Analysed: 50 Age: 28 years Gender: 78% men Recruited: from Rheumatism Center at the University Medical Center and home, monocentre Inclusion criteria: patients with AS (a) an outpatient without complications, (b) sedentary, as defined by a lack of regular exercise during the previous 6 months, (c) able to understand the content of questionnaires and experimental schedules, (d) had no changes in their current prescription medication, and (e) were classified in the functional class II for AS Exclusion criteria: not reported

Lim 2005 *(Continued)*

	Severity of the disease: classified as functional class II Disease duration: 9 years Coexisting medication treatment: no change of medication during the study
Interventions	Exercise group (N = 25) • Monomodal programme with exercise • Exercise components: "Sixteen movements based on the exercise programme recommended by the Spondylitis Association of America". Muscle relaxation, flexibility, muscular strength, stronger breathing, and straight posture: ‘‘stretch out’’, ‘‘cat-back’’ (sway-back), ‘‘hands and knees rock’’, neck flexion and extension, neck lateral move-ment, body rotation, hip flexor–quadriceps stretch,hamstring stretch, abdominal strengthening, hip extensor exercise, alternative hip extensor exercise,breathing, ‘‘shoulder circle’’, and pectoral muscle stretch. • Dose: 30 min/day * 8 weeks duration • Equipment: Not reported • Delivery mode: self delivery, patients receiving exercises were asked to practice these exercises at home individually for 8 weeks and were telephoned by the researchers every day. • Provider: by "an expert and a researcher taught the exercise motions". • Tailoring: not reported Control group (N = 25) • No intervention • Participants were on a waiting list Adherence • Compliance: no information reported • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0, final point = 8 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Pain: VAS (0 to 100) • Safety: not reported as an outcome Minor outcomes • Not reported
Notes	Unclear if participants in the control group could practice community-based exercises Number of missing participants: 8 (14%) Dropouts or withdrawals: n = 0 Adverse events: n = 0 (0%) Averse effects: not reported Funding source: not reported Protocol number registration not found Declaration of interest: not reported
Risk of bias	

Lim 2005 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Comment: no clear information provided in the report Quote: "Subjects were randomly assigned to either an exercise group or a wait-list control group."
Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: participants were not blinded Quote: "The subjects were informed about the exercise therapy when we explained the nature of AS and procedures for the study."
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: no intention-to-treat analysis; 8 dropouts; no imputation technique was used
Selective reporting (reporting bias)	Unclear risk	Comment: no registration protocol reported; outcomes listed in the method section were all reported in the results section; no major outcomes specified
Other bias	Unclear risk	Comment: small number of subjects; unclear description of the intervention; no power sample calculation; multiple outcomes from comparisons with no inflation risk alpha correction; funding and conflict of interest were not declared; attendance and compliance were not reported

Masiero 2011

Methods	RCT with three groups No intention-to-treat analysis
Participants	Location: Italy Randomised: 45 Analysed: 42 Age: 47 years Gender: 79% men Recuited: from outpatient of a rheumatology hospital department Inclusion criteria: treated by anti-TNF medication (infliximad, etanercept, or adalimumad) for a least 9 months, no major change of the clinical status in the last 3 months, aged between 18 and 65 years, diagnosis of AS on the modified New York criteria, no severe disease associated Exclusion criteria: suffering from complete spine ankylosing, being involved in other rehabilitation treatments, failure to take part in the study, variations in standard biological therapy regimens during the study Severity and duration of the disease: the mean disease duration was 9 years

Masiero 2011 *(Continued)*

	Coexisting medication treatment: NSAIDS (0%)
Interventions	Exercise group • Multimodal programme with exercise, education meetings • Exercise components: flexibility, muscle stretching, proprioceptive training, and exercise to expand the chest and control breathing. Participants were encouraged to practice at home • Dose: 60 min sessions, twice weekly, for 12 weeks • Equipment: at the end of each session, participants received a brochure with a home guide • Delivery mode: interdisciplinary team; group sessions under supervision and home delivery. Setting: rehabilitation unit and home • Provider: by an experienced physiotherapist • Tailoring: at the start of each session, feedback was given and problems with home practice were discussed Control group • No intervention group • "Participants received no rehabilitation" • Limited description Adherence • Compliance: not sufficient information reported • Attendance: not sufficient information reported
Outcomes	Time points (weeks): baseline = 0, final point = 8, follow-up = 16 Major outcomes • Pain: cervical and lumbar pain (VAS 0 to 10) • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Spinal mobility: cervical rotation (degrees), chest expansion (in cm), BASMI (0 to 10 scale) • Fatigue: BASDAI fatigue (0 to 10) • Safety: not reported as an outcome Minor outcomes
Notes	Unclear if participants in the control group could practice community-based exercises Number of missing participants: 3 (7%) Dropouts or withdrawals: n = 3 (7%) two withdraws and one dropout Adverse events: not reported Averse effects: not reported Funding source: not reported Registration number: protocol not found Declaration of interest: not reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Comment: no clear information provided in the report

Masiero 2011 *(Continued)*

		Quote: "Subjects were randomly allocated to attend either rehabilitation therapy."
Allocation concealment (selection bias)	Low risk	Quote: "Casual randomisation using a statistical programme was carried out by a rheumatologist not involved in the study evaluation or rehabilitation intervention"
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: not described, but it's unlikely that personnel were blinded to the treatment allocated
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: few balanced dropouts (2 and 1 in the groups). May have a minor influence on the estimates effect size
Selective reporting (reporting bias)	Unclear risk	Comment: outcomes listed in the method section were all reported in the results section. No protocol registration was found. No major outcomes specified
Other bias	Unclear risk	Comment: attendance and compliance were not reported; small sample size; baseline performances were similar for major outcomes; no power sample calculation; funding and conflict of interest were not stated

Rodriguez-Lozano 2013

Methods	RCT with two groups No intention-to-treat analysis
Participants	Location: Spain Randomised: 802 Analysed: 756 Age: 45 years Gender: 72% men Recuited: from outpatient of rheumatology services, multicentre (24 hospitals) Inclusion criteria: aged 18 to 70 years, diagnosis of AS on the modified New York criteria Exclusion criteria: suffered from severe form of AS with loss of motion, form of AS or coexistent disease with a contraindication for exercises Severity and duration of the disease: the mean duration of disease was 17 years, low to moderate disease activity Coexisting medication treatment: NSAIDS (75%), corticosteroids (4%), biologic agents (39%), analgesics (11.5%), Sulfasalazine (9%)
Interventions	Exercise group (N = 381)

Rodriguez-Lozano 2013 *(Continued)*

- Multimodal programme with exercise, education and psychological support
- Exercise components: stretching, breathing, active joint motion with exercise recommendation to practice at home
- Dose: 60 min sessions, 7/week, for 24 weeks
- Equipment: participation of one of the family members, DVD and booklet of the programme to take home
- Delivery mode: group sessions and home delivery
- Provider: exercises developed by a rehabilitation specialist and a physiotherapist and self delivery. Setting: unclear
- Tailoring: on-site practice session to carry out the most difficult exercises with the help of the physiotherapist

Control group (N = 375)

- usual care with pharmacological and non-pharmacological treatments

Adherence

- Compliance: weekly diary for the number of exercise performed and use of NSAIDs
- Attendance: no information reported

Outcomes	Time points (weeks): baseline = 0, final point = 24 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Pain: VAS total pain and nocturnal pain (VAS 0 to 10) • Safety: not reported as an outcome Minor outcomes: • Quality of life: ASQoL (0 to 18)
Notes	Number of missing participants: 57 (7%) Dropouts or withdrawals: n = 46 (6%) "failed to attend the final visit" Adverse events: not reported Averse effects: not reported Funding source: from the Spanish Society of Rheumatology Registration number: protocol not found Declaration of interest: authors declared no conflict of interest

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "The central agency was located in Madrid. A random sequence for the allocation was generated. It was sent to every hospital in an opaque envelope".
Allocation concealment (selection bias)	Low risk	Comment: sealed envelopes with assignment code were used Quote: "concealment of allocation was assured by opening an opaque envelope which contained the assignation number"... "contained a consecutive number".

Rodriguez-Lozano 2013 *(Continued)*

Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: no intention-to-treat analysis; missing data; no imputation technique was used; possible influence on treatment effect
Selective reporting (reporting bias)	Unclear risk	Comment: outcomes listed in the method section were all reported in the results section; no protocol registration was found; no major outcomes specified
Other bias	Low risk	Comment: convenient number of subjects included in the study; adherence to treatments was monitored; power sample size calculation was reported; size effect calculation adjusted on baseline variables

Souza 2017

Methods	RCT with two groups intention-to-treat analysis
Participants	Location: Brazil Randomised: 60 Analysed: 60 Age: 44 years Gender: 73% men Recruited: outpatient clinic of a university hospital Inclusion criteria: diagnosis of AS on the modified New York criteria, a Steinbrocker functional class of I to II, on stable medication, disease-modifying anti-rheumatic drug (DMARDS) for at least 3 months, and NSAIDs or corticoids (or both) for at least 4 weeks. If not using medication for AS, should be medication-free for at least 3 months Exclusion criteria: uncontrolled hypertension, history of coronary artery disease, history of syncope or arrhythmias induced by exercise, decompensated diabetes mellitus, severe psychiatric disorders, fibromyalgia, a more disabling condition than AS, a history of regular exercise of at least 30 min, two times a week, in the last 3 months, any condition that could prevent the patient from performing exercises in the last three months Severity and duration of the disease: a Steinbrocker functional class of I to II; mean time since diagnosis of 9.2 years Coexisting medication treatment: NSAIDS (27%), biologic agents (38%), DMARDS (17%), Sulfasalazine (12%), no medication (8%)
Interventions	Exercise group (N = 27) • Monomodal programme with exercises • Exercise components: strengthening (8 exercises); exercises performed with 3 sets of 10 repetitions each

Souza 2017 *(Continued)*

- Dose: 50 min, 2 days a week, for 16 weeks
- Equipment: Swiss ball
- Delivery mode: group session supervised; setting not reported
- Provider: by a trained physiotherapist.
- Tailoring: ball size according to patient height; weight progression based on 1 RM assessment (50% at beginning, 60% after 4 weeks, and 70% after 12 weeks)

Control group (N = 28)

- No intervention group
- "control group remained with medical treatment only"

Adherence

- Compliance: no information reported
- Attendance: no information reported

Outcomes	Time points (weeks): baseline = 0, final point = 16 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Pain: SF-36 • HAQ-S questionnaire (range 0 to 3) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Spinal mobility: BASMI (VAS 10-cm scale), • Chest expansion (in cm) • Safety: not reported as an outcome Minor outcomes • Quality of life: SF-36 (0 to 100) • ESR (mm/h) • CRP (mg/dL)
Notes	Number of missing participants: none Dropouts or withdrawals: 5 participants dropped out Adverse events: not explicitly reported Averse effects: not explicitly reported Sample size calculation: 27 subjects per arm Funding source: granted by a research foundation (# 2011/03459-9) Declaration of interest: authors declared no conflict Protocol registration number: NCT0351311

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "a computer-generated randomisation list was utilised to randomly allocate patients into intervention (IG) or control (CG) groups".
Allocation concealment (selection bias)	Unclear risk	incomplete information provided

Souza 2017 *(Continued)*

		Quote: "a concealed randomization with an opaque, sealed envelope"
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded Quote: "The evaluations were performed by a blinded evaluator "
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: intention-to-treat analysis; technique of imputation – the last evaluation carried forward; reasons for dropouts were not clearly reported
Selective reporting (reporting bias)	Low risk	Comment: protocol number registration was provided; no difference was found between the protocol and the published report; all the outcomes prespecified were reported
Other bias	Unclear risk	Comment: attendance and compliance not reported; no systematic reporting of adverse events; no outcomes reported at medium- and long-term follow-up; power sample size calculation; no baseline adjustment of data

Sveaas 2014

Methods	RCT with two groups No intention-to-treat
Participants	Location: Norway Randomised: 28 Analysed: 24 Age: 48 years Gender: 50% men Recruited: from Diakonhjemmet Hospital, monocentre Inclusion criteria: axSpA according to the Assessment of SpA International Society (ASAS) classification criteria; age 18 to 70 years; no change in TNF inhibitor use during the last 3 months, moderate to high disease activity (Bath AS Disease Activity Index (BASDAI)), did not perform regular endurance or strength exercise during the last year (1 hour per week) Exclusion criteria: established CVD, other comorbidity involving reduced exercise capacity, inability to participate in weekly exercise sessions in Oslo, pregnancy Severity and duration of the disease: mean ASDAS = 2.6, mean duration = 25 years Coexisting medication treatment: NSAIDS (75%), TNF inhibitor (29%)
Interventions	Exercise group (N =13) • Monomodal programme with exercise • Exercise components: endurance and strength training of 20 minutes of major muscles group • Dose: 40 to 60 min, 3/week, for 12 weeks; endurance: high intensity interval training on a treadmill for 40 minutes (90% to 95% of maximum heart rate)

Sveaas 2014 *(Continued)*

	• Equipment: bench press, rowing • Delivery mode: at a fitness centre with individual supervision twice a week, and once a week individually • Provider: physical therapist • Tailoring: exercise programme intensity individually adapted Control group (N = 15) • No intervention group • "Participants in the CG were asked to not start exercising during the intervention period" Adherence • Compliance: attendance was recorded by the physiotherapist. The participants had to follow at least 80% of the planned exercise sessions to fulfil the exercise protocol • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0, final point = 12 Major outcomes • Physical function: BASFI (NRS 0 to 10) • Patient global assessment of disease activity: BASDAI (NRS 0 to 10) • Spinal mobility: BASMI (0 to 10 scale) • Safety: any adverse events were reported as an outcome Minor outcomes • Acute-phase reactant CRP level (mg/L) • ESR
Notes	Number of missing participant: 4 (14%) Dropouts or withdrawals: n = 4 (14%) Adverse events: n = 3 (10%) in the exercise group Averse effects: n = 1 (4%) in the exercise group, because intervention was physically challenging and time consuming Funding source: by the Norwegian Foundation for Postgraduate Physiotherapists Protocol registration number: NCT01436942 Declaration of interest: authors declared they had no competing interests

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "Allocation to EG or CG followed a computer-generated randomisation list"
Allocation concealment (selection bias)	Unclear risk	Comment: incomplete description of the allocation procedure Quote: the group assignment was concealed in numbered envelopes, and revealed consecutively after baseline testing
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated

Sveaas 2014 *(Continued)*

Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded Quote: "The assessors were blinded for group assignment".
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: per-protocol analysis; no imputation technique was used; possible influence on treatment effect
Selective reporting (reporting bias)	High risk	Comment: a protocol registration number was provided. Different outcomes were mentioned in the protocol (generic General Health Questionnaire (GHQ-12); international Physical Activity Questionnaire short version (IPAQ-s), which were not reported in the published article. There is a possible risk of bias.
Other bias	Unclear risk	Comment: limited number of subjects per group; size effect estimates adjusted for baseline performance (covariance analysis); no power sample calculation; multiple outcomes for comparisons with no inflation risk alpha correction

Sweeney 2002

Methods	RCT with two groups No intention-to-treat analysis
Participants	Location: United Kingdom Randomised: 200 Analysed: 155 Age: 47 years Gender: 69% men Recruited: from members of outpatient service or National Ankylosing Spondylitis Society (NASS) Inclusion criteria: 16 to 65 years, outpatients of the RNHRD or members of the National Ankylosing Spondylitis Society (NASS) Exclusion criteria: not reported Severity and duration of the disease: the duration of disease was 21 years Coexisting medication treatment: non reported
Interventions	Exercise group (N = 75) • Multidisciplinary programme with exercise and education • Exercise components: incomplete description • Dose: incomplete description • Equipment: video • Delivery mode: sent by mail; included video, booklet, and reminder stickers. Setting: exercises performed at home • Provider: self-delivery • Tailoring: not described

Sweeney 2002 *(Continued)*

	Control group (N = 80) • description unclear • authors mentioned "standard care patients" in their report Adherence • Compliance: no information reported • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0, final point = 24 Major outcomes • Physical function: BASFI (0 to 10 scale) • Patient global assessment of disease activity: BASDAI (0 to 10 scale) • Pain: Stanford Self-efficacy scale. The score is the mean of five items; each items is expressed on a scale from 0 to 10 (lower score is worse) • Safety: not reported as an outcome Minor outcomes • Not reported
Notes	Number of missing participant: 45 (22%) Dropouts or withdrawals: n = 45 Adverse events: not reported Averse effects: not reported Funding source: supported by grants from BUPA, National Ankylosing Spondylitis Society, John Coates Charitable Trust, and Col. W.W. Pilkington Trust Protocol registration number not found Declaration of interest: not reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Comment: incomplete description Quote: "The selected patients were then randomly assigned to 2 groups".
Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: unsupervised home delivery; personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: data were missing; no ITT analysis; can have a possible influence on the estimates of effect size

Sweeney 2002 *(Continued)*

Selective reporting (reporting bias)	Unclear risk	Comment: no registration protocol number reported; major and minor outcomes were not specified; outcomes listed in the methods section were all reported in the results section; reasons for dropouts were not reported
Other bias	Unclear risk	Comment: no sample size calculation; sample size > 30; multiple statistical comparisons without corrections; unclear if it could have an influence on the estimates of effect size; adverse effects were not monitored or reported; funding and conflict of interest were not stated

Widberg 2009

Methods	RCT with 2 groups intention-to-treat analysis
Participants	Location: Sweden Randomised: 32 Analysed: 32 Age: 36 years Gender: 100% men Recruited: at Karolinska University Hospital, monocentre Inclusion criteria: diagnosis of ankylosing spondylitis according to the modified New York criteria, 20 to 60 years old, and had stable pharmacological treatment with non-steroidal anti-inflammatory drugs and disease-modifying anti-rheumatic drugs Exclusion criteria: inflammatory disease activity, which could be subject to pharmacological changes, radiological ossification between the thoracic vertebrae, concomitant effects of other severe illnesses Severity and duration of the disease: duration of disease (median) = 2.5 to 3.5 years Coexisting medication treatment: NSAIDS (75%), and DMARDS (44%)
Interventions	Exercise group (N = 16) • Monomodal programme with exercise • Exercise components: passive and active joint mobility with stretching of tight muscles, home exercise; warming up the soft tissue of the back muscles (with vibrations via a vibrator) and gentle mobility exercises. Active angular and passive mobility exercises in the physiological directions of the joints in the spinal column and in the chest wall in three directions of motion (flexion/extension, lateral flexion, and rotation), and in different starting positions (lying face down, sideways, on the back, and in a sitting position). Passive mobility exercises consisted of general, angular movements and specific, translatory movements. • Dose: 60 min, 2/week, for 8 weeks • Equipment: not reported • Delivery mode: supervised; guided by the patient's current disease activity, balancing between increasing pain and yet improving mobility; setting: outpatient and home • Provider: delivered by a physiotherapist • Tailoring: exercises were adjusted by participant's pain level; unclear how they were adapted Control group (N = 16) • Usual care group • patients were encouraged to perform their usual physical exercises during the eight weeks

Widberg 2009 *(Continued)*

	Adherence • Compliance: asked at the physiotherapist visit; all patients participated in all sessions • Attendance: no information reported
Outcomes	Time points (weeks): baseline = 0, final point = 8 Major outcomes • Physical function: BASFI (VAS 10-cm scale) • Patient global assessment of disease activity: BASDAI (VAS 10-cm scale) • Spinal mobility: BASMI (VAS 10-cm scale); chest expansion (cm) • Safety: not reported as an outcome Minor outcomes
Notes	Number of missing participants: 0 (0%) Dropouts or withdrawals: n = 0 Adverse events: not reported Averse effects: not reported Funding source: Swedish Rheumatism Association and Nacka Rehab centre Protocol registration number not found Declaration of interest: authors declared no conflict of interest

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Comment: adequate method of random sequence generation Quote: "The participants were assigned randomly into two groups using random number table by the researcher other than the one who performed the evaluation throughout the study."
Allocation concealment (selection bias)	Unclear risk	Comment: the study did not provide information about allocation concealment
Blinding of participants (subjective outcome)	High risk	Comment: not described, but it's unlikely that participants were blinded to the treatment allocated
Blinding of personnel	High risk	Comment: personnel unblinded
Blinding of outcome assessment (subjective)	High risk	Comment: participants were the assessors; they were unblinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: all the patients randomised were analysed
Selective reporting (reporting bias)	Unclear risk	Comment: no protocol registration number reported; outcomes listed in the methods section were all reported in the results section; safety was not reported as an outcome; major outcomes were specified

Widberg 2009 *(Continued)*

Other bias	Unclear risk	Comment: small sample size; baseline performances were similar for major outcomes; rates of medications did not differ between the two groups; no power sample calculation

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Ciprian 2013	Does not meet the intervention inclusion criteria. The exercise programme was combined with different cointerventions: thermal water and mud pack
Colina 2009	Controlled study with no randomisation
Durmus 2009	Controlled study with no randomisation
Gunay 2012	A quasi-randomised study: all patients were randomised with their outpatient clinic registration numbers
Karahan 2016	Does not meet the intervention inclusion criteria. The experimental intervention included recreational physical activities
Lee 2008	Controlled study with no randomisation
Masiero 2015	Controlled study with no randomisation

Characteristics of studies awaiting assessment *[ordered by study ID]*

Mesquita 2014

Methods	RCT
Participants	Patients with ankylosing spondylitis
Interventions	Home-based exercise programme; no information on control intervention
Outcomes	BASDAI, BASFI, BASMI, WHOQOL-bref; measurements at baseline and 10 weeks after
Notes	Abstract of congress; incomplete data; we contacted the authors but they did not respond

Sveeas 2018

Methods	RCT
Participants	Patients with axial spondyloarthritis
Interventions	Exercise group (EG), which performed cardiorespiratory and strength exercises, or a control group (CG), which received treatment as usual
Outcomes	Emotional distress, fatigue, and ability to do a full day's activities, measurements at baseline and 12 weeks after

Sveeas 2018 *(Continued)*

Notes	Same authors as Sveaas 2014; it might be the same study with different outcomes; awaiting confirmation from authors

Characteristics of ongoing studies *[ordered by study ID]*

ChiCTR-TRC-14004650

Trial name or title	The effect of traditional exercise 'Baduanjin' for physical functioning of ankylosing spondylitis: a randomised, controlled, prospective study
Methods	Randomised parallel control
Participants	experimental group: 30; control group: 30
Interventions	experimental group: Baduanjin exercise; control group: maintain present treatment and lifestyle
Outcomes	BASDAI index; BASMI index; BASFI index; HAQ-9; patients global
Starting date	16 May 2014
Contact information	77612802@qq.com
Notes	Study recruiting; apps.who.int/trialsearch/Trial2.aspx?TrialID=ChiCTR-TRC-14004650

NCT02098694

Trial name or title	Physiotherapy-led outpatient clinic for patients with spondyloarthritis
Methods	RCT
Participants	outpatient with spondyloarthritis
Interventions	home-exercises (experimental) versus usual care (control)
Outcomes	BASMI, BASFI, ASDAS. measurements at baseline and at 8 months
Starting date	June 2014
Contact information	Ann-Katrin Stensdotter, Norwegian University of Science and Technology
Notes	Study completed. NCT02098694

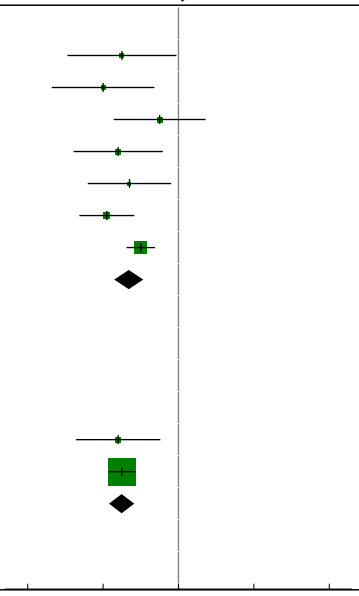
DATA AND ANALYSES

Comparison 1. Exercise vs no intervention

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Physical function	7		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 BASFI at end of intervention	7	312	Mean Difference (IV, Random, 95% CI)	-1.32 [-1.71, -0.93]
1.2 BASFI at medium-term follow-up	2	93	Mean Difference (IV, Random, 95% CI)	-1.51 [-1.84, -1.17]
2 Pain	6		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 End of intervention	6	288	Std. Mean Difference (IV, Random, 95% CI)	-0.82 [-1.40, -0.25]
2.2 Pain at medium term follow-up	2	93	Std. Mean Difference (IV, Random, 95% CI)	-2.50 [-5.32, 0.32]
3 Patient global assessment of disease activity	6		Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 BASDAI at end of intervention	6	262	Mean Difference (IV, Random, 95% CI)	-0.91 [-1.32, -0.49]
3.2 BASDAI at medium-term follow-up	2	93	Mean Difference (IV, Random, 95% CI)	-1.12 [-1.57, -0.68]
4 Spinal mobility	5		Mean Difference (IV, Random, 95% CI)	Subtotals only
4.1 BASMI at end of intervention	5	232	Mean Difference (IV, Random, 95% CI)	-0.70 [-1.28, -0.13]
4.2 BASMI at medium-term follow-up	2	93	Mean Difference (IV, Random, 95% CI)	-1.42 [-2.05, -0.78]
5 Fatigue	2	72	Mean Difference (IV, Random, 95% CI)	-1.43 [-2.73, -0.14]
5.1 BASDAI at end of intervention	2	72	Mean Difference (IV, Random, 95% CI)	-1.43 [-2.73, -0.14]
6 Quality of life	2	85	Mean Difference (IV, Random, 95% CI)	1.74 [-0.44, 3.91]
6.1 QQL at end of intervention	2	85	Mean Difference (IV, Random, 95% CI)	1.74 [-0.44, 3.91]
7 C-Reactive Protein (CRP)	2	84	Mean Difference (IV, Random, 95% CI)	1.38 [-4.34, 7.10]
8 Erythrocyte Sedimentation Rate (ESR)	2	84	Mean Difference (IV, Random, 95% CI)	-5.36 [-10.31, -0.41]

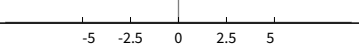
Analysis 1.1. Comparison 1 Exercise vs no intervention, Outcome 1 Physical function.

Study or subgroup	Exercise N	Exercise Mean(SD)	No intervention N	No intervention Mean(SD)	Mean Difference Random, 95% CI	Weight	Mean Difference Random, 95% CI
1.1.1 BASFI at end of intervention							
Gallinaro 2016	37	3 (2.7)	18	4.5 (2.5)		6.4%	-1.5[-2.95,-0.05]
Garcia 2015	15	3.8 (1.8)	15	5.8 (2)		7.12%	-2[-3.36,-0.64]
Souza 2017	30	3.4 (2.2)	30	3.9 (2.6)		8.64%	-0.5[-1.72,0.72]
Sveaas 2014	10	1.5 (1.5)	14	3.1 (1.4)		9.08%	-1.6[-2.78,-0.42]
Dönmez 2014	25	1.7 (1.4)	26	3 (2.5)		10.19%	-1.3[-2.41,-0.19]
Lim 2005	25	1.6 (1.1)	25	3.5 (1.5)		19.38%	-1.9[-2.63,-1.17]
Masiero 2011	20	1.8 (0.6)	22	2.8 (0.7)		39.18%	-1[-1.38,-0.62]
Subtotal ***	**162**		**150**			**100%**	**-1.32[-1.71,-0.93]**
Heterogeneity: Tau²=0.06; Chi²=7.79, df=6(P=0.25); I²=22.96%							
Test for overall effect: Z=6.7(P<0.0001)							
1.1.2 BASFI at medium-term follow-up							
Dönmez 2014	25	1.4 (1.2)	26	3 (2.7)		8.97%	-1.6[-2.72,-0.48]
Masiero 2011	20	1.2 (0.4)	22	2.7 (0.7)		91.03%	-1.5[-1.85,-1.15]
Subtotal ***	**45**		**48**			**100%**	**-1.51[-1.84,-1.17]**
Heterogeneity: Tau²=0; Chi²=0.03, df=1(P=0.87); I²=0%							
Test for overall effect: Z=8.83(P<0.0001)							



Analysis 1.2. Comparison 1 Exercise vs no intervention, Outcome 2 Pain.

Study or subgroup	Exercise N	Exercise Mean(SD)	No intervention N	No intervention Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
1.2.1 End of intervention							
Garcia 2015	15	5.1 (1.7)	15	7 (1.8)		15.1%	-1.08[-1.85,-0.31]
Masiero 2011	20	1.5 (0.7)	22	2.7 (1)		16.07%	-1.41[-2.09,-0.73]
Lim 2005	25	3.1 (1.7)	25	5.6 (1.3)		16.43%	-1.66[-2.31,-1.01]
Dönmez 2014	25	1.6 (1.7)	26	3.5 (2.4)		17.18%	-0.89[-1.47,-0.31]
Gallinaro 2016	37	2.2 (1.8)	18	2.3 (2.7)		17.33%	-0.06[-0.62,0.51]
Souza 2017	30	3.4 (1.9)	30	3.4 (2.8)		17.9%	0[-0.51,0.51]
Subtotal ***	**152**		**136**			**100%**	**-0.82[-1.4,-0.25]**
Heterogeneity: Tau²=0.41; Chi²=26.18, df=5(P<0.0001); I²=80.9%							
Test for overall effect: Z=2.81(P=0)							
1.2.2 Pain at medium term follow-up							
Masiero 2011	20	0.9 (0.5)	22	3 (0.6)		48.72%	-3.98[-5.05,-2.9]
Dönmez 2014	25	1.2 (1.4)	26	3.4 (2.4)		51.28%	-1.1[-1.69,-0.5]
Subtotal ***	**45**		**48**			**100%**	**-2.5[-5.32,0.32]**
Heterogeneity: Tau²=3.95; Chi²=21.02, df=1(P<0.0001); I²=95.24%							
Test for overall effect: Z=1.74(P=0.08)							
Test for subgroup differences: Chi²=1.3, df=1 (P=0.25), I²=23.08%							



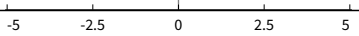
Analysis 1.3. Comparison 1 Exercise vs no intervention, Outcome 3 Patient global assessment of disease activity.

Study or subgroup	Exercise N	Exercise Mean(SD)	No intervention N	No intervention Mean(SD)	Mean Difference Random, 95% CI	Weight	Mean Difference Random, 95% CI
1.3.1 BASDAI at end of intervention							
Sveaas 2014	10	3.3 (2)	14	5.2 (2)		6.05%	-1.9[-3.52,-0.28]
Dönmez 2014	25	2 (1.2)	26	3.5 (2.7)		11.49%	-1.5[-2.64,-0.36]
Gallinaro 2016	37	2.3 (1.8)	18	3.2 (2)		12.39%	-0.95[-2.04,0.14]
Souza 2017	30	2.1 (1.8)	30	2.1 (2.4)		12.56%	-0.04[-1.12,1.04]
Garcia 2015	15	2.8 (1.2)	15	4 (1.3)		17.93%	-1.22[-2.09,-0.35]
Masiero 2011	20	2.3 (0.7)	22	3 (0.9)		39.59%	-0.7[-1.19,-0.21]
Subtotal ***	**137**		**125**			**100%**	**-0.91[-1.32,-0.49]**
Heterogeneity: Tau^2=0.05; Chi^2=6.07, df=5(P=0.3); I^2=17.65%							
Test for overall effect: Z=4.3(P<0.0001)							
1.3.2 BASDAI at medium-term follow-up							
Dönmez 2014	25	2 (1.3)	26	3.2 (2.6)		15.77%	-1.25[-2.37,-0.13]
Masiero 2011	20	2.1 (0.7)	22	3.2 (0.9)		84.23%	-1.1[-1.59,-0.61]
Subtotal ***	**45**		**48**			**100%**	**-1.12[-1.57,-0.68]**
Heterogeneity: Tau^2=0; Chi^2=0.06, df=1(P=0.81); I^2=0%							
Test for overall effect: Z=4.94(P<0.0001)							
Test for subgroup differences: Chi^2=0.49, df=1 (P=0.48), I^2=0%							

Favours exercise — Favours no intervention

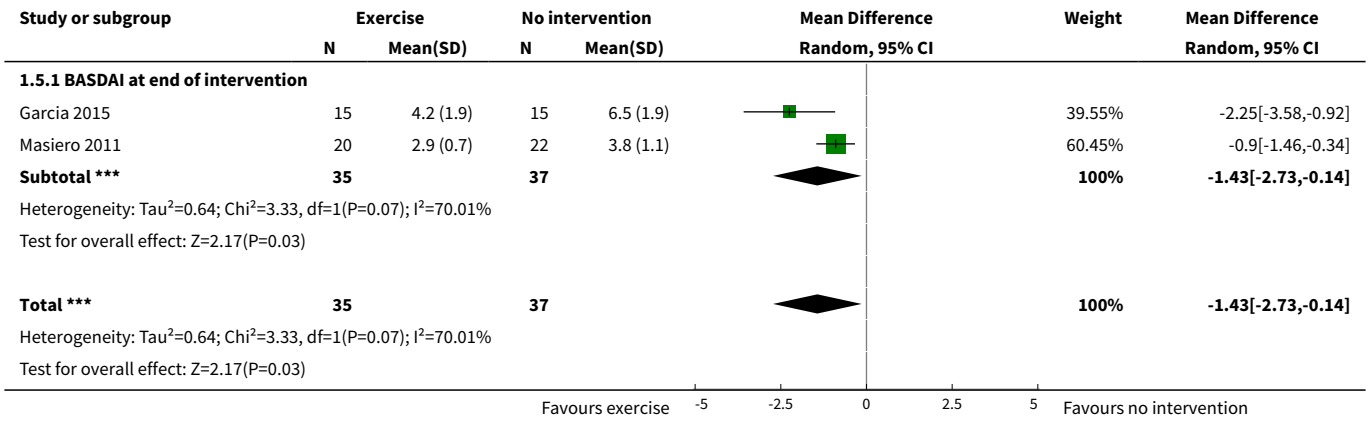
Analysis 1.4. Comparison 1 Exercise vs no intervention, Outcome 4 Spinal mobility.

Study or subgroup	Exercise N	Exercise Mean(SD)	No intervention N	No intervention Mean(SD)	Mean Difference Random, 95% CI	Weight	Mean Difference Random, 95% CI
1.4.1 BASMI at end of intervention							
Sveaas 2014	10	2 (1.6)	14	2.9 (1.8)		12.28%	-0.9[-2.27,0.47]
Gallinaro 2016	37	4.4 (2.2)	18	4.6 (2)		15.25%	-0.2[-1.36,0.96]
Souza 2017	30	4.7 (1.9)	30	5.4 (2.2)		17.5%	-0.7[-1.74,0.34]
Dönmez 2014	25	2.9 (1.7)	26	4.6 (1.7)		19.75%	-1.7[-2.63,-0.77]
Masiero 2011	20	3.7 (0.6)	22	4 (0.7)		35.22%	-0.3[-0.68,0.08]
Subtotal ***	**122**		**110**			**100%**	**-0.7[-1.28,-0.13]**
Heterogeneity: Tau^2=0.2; Chi^2=8.1, df=4(P=0.09); I^2=50.63%							
Test for overall effect: Z=2.42(P=0.02)							
1.4.2 BASMI at medium-term follow-up							
Dönmez 2014	25	2.6 (1.7)	26	4.5 (1.7)		31.27%	-1.9[-2.83,-0.97]
Masiero 2011	20	3.1 (0.4)	22	4.3 (0.9)		68.73%	-1.2[-1.61,-0.79]
Subtotal ***	**45**		**48**			**100%**	**-1.42[-2.05,-0.78]**
Heterogeneity: Tau^2=0.11; Chi^2=1.82, df=1(P=0.18); I^2=44.91%							
Test for overall effect: Z=4.37(P<0.0001)							
Test for subgroup differences: Chi^2=2.68, df=1 (P=0.1), I^2=62.74%							

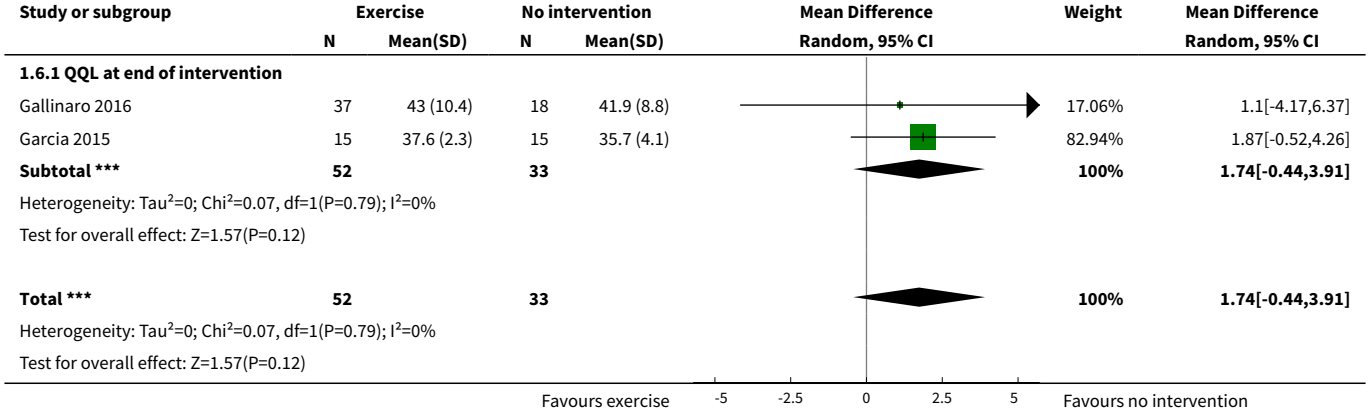


Favours exercise — Favours no intervention

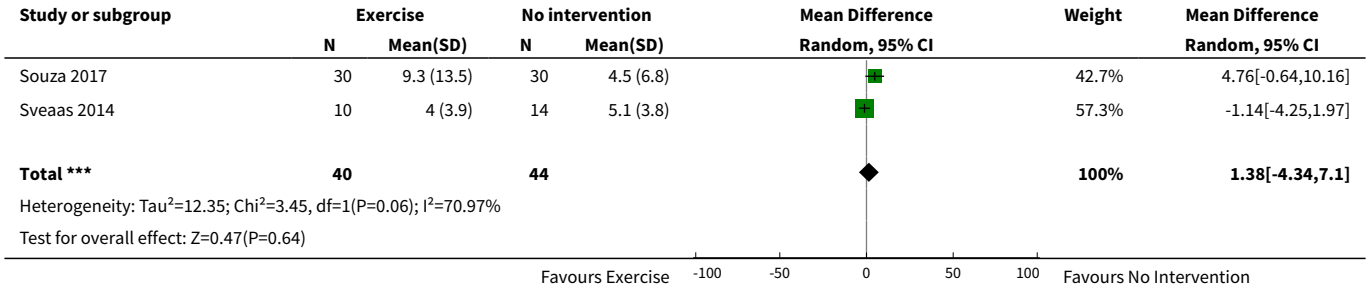
Analysis 1.5. Comparison 1 Exercise vs no intervention, Outcome 5 Fatigue.



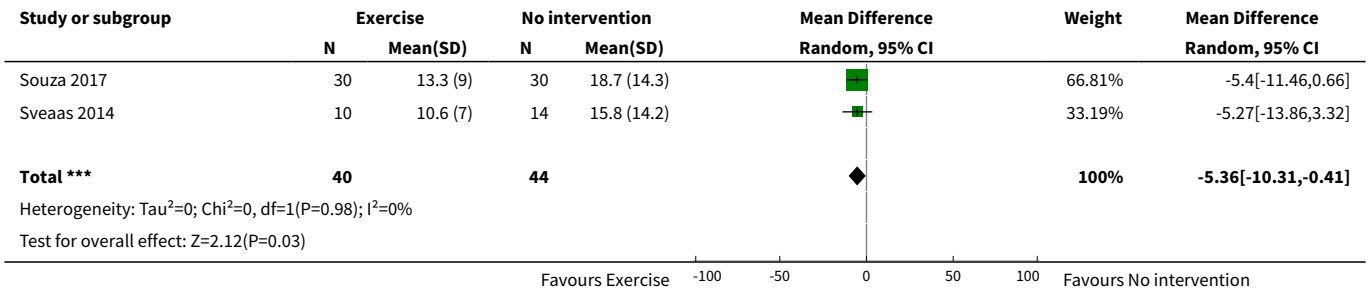
Analysis 1.6. Comparison 1 Exercise vs no intervention, Outcome 6 Quality of life.



Analysis 1.7. Comparison 1 Exercise vs no intervention, Outcome 7 C-Reactive Protein (CRP).



Analysis 1.8. Comparison 1 Exercise vs no intervention, Outcome 8 Erythrocyte Sedimentation Rate (ESR).

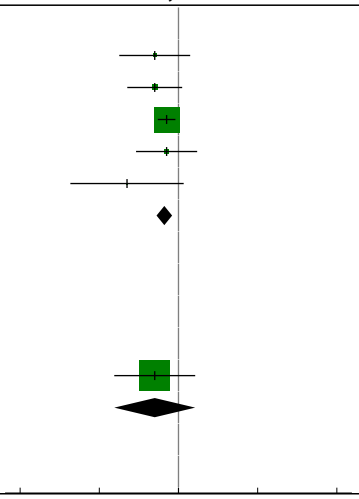


Comparison 2. Exercise vs usual care

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Physical function	5		Mean Difference (IV, Fixed, 95% CI)	Subtotals only
1.1 BASFI at end of intervention	5	1068	Mean Difference (IV, Fixed, 95% CI)	-0.36 [-0.55, -0.16]
1.2 BASFI at medium-term follow-up	1	53	Mean Difference (IV, Fixed, 95% CI)	-0.60 [-1.62, 0.42]
2 Pain	2		Std. Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 End of intervention	2	911	Std. Mean Difference (IV, Random, 95% CI)	-0.16 [-0.29, -0.03]
3 Patient global assessment of disease activity	5		Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 BASDAI at end of intervention	5	1068	Mean Difference (IV, Random, 95% CI)	-0.68 [-1.27, -0.09]
3.2 BASDAI at medium-term follow-up	1	53	Mean Difference (IV, Random, 95% CI)	-0.70 [-1.71, 0.31]
4 Spinal mobility	2		Mean Difference (IV, Random, 95% CI)	Subtotals only
4.1 BASMI at end of intervention	2	85	Mean Difference (IV, Random, 95% CI)	-1.15 [-2.81, 0.52]
4.2 BASMI at medium-term follow-up	1	53	Mean Difference (IV, Random, 95% CI)	-0.70 [-1.64, 0.24]
5 Quality of life	2		Mean Difference (IV, Random, 95% CI)	Subtotals only
5.1 QQL at end of intervention	2	809	Mean Difference (IV, Random, 95% CI)	-0.36 [-1.68, 0.95]

Analysis 2.1. Comparison 2 Exercise vs usual care, Outcome 1 Physical function.

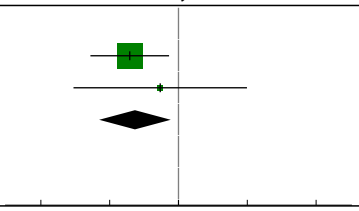
Study or subgroup	Exercise N	Exercise Mean(SD)	Usual care N	Usual care Mean(SD)	Mean Difference Fixed, 95% CI	Weight	Mean Difference Fixed, 95% CI
2.1.1 BASFI at end of intervention							
Altan 2012	29	1.7 (1.6)	24	2.3 (1.7)		4.84%	-0.6[-1.5,0.3]
Kjeken 2013	37	3.4 (1.5)	35	4 (1.5)		8.08%	-0.6[-1.29,0.09]
Rodriguez-Lozano 2013	381	-0.5 (1.4)	375	-0.2 (1.7)		78.66%	-0.3[-0.52,-0.08]
Sweeney 2002	75	3.1 (2.3)	80	3.4 (2.6)		6.52%	-0.3[-1.07,0.47]
Widberg 2009	16	2 (1.8)	16	3.3 (2.3)		1.9%	-1.3[-2.73,0.13]
Subtotal ***	**538**		**530**			**100%**	**-0.36[-0.55,-0.16]**
Heterogeneity: Tau^2=0; Chi^2=2.7, df=4(P=0.61); I^2=0%							
Test for overall effect: Z=3.56(P=0)							
2.1.2 BASFI at medium-term follow-up							
Altan 2012	29	1.7 (1.6)	24	2.3 (2.1)		100%	-0.6[-1.62,0.42]
Subtotal ***	**29**		**24**			**100%**	**-0.6[-1.62,0.42]**
Heterogeneity: Not applicable							
Test for overall effect: Z=1.15(P=0.25)							



Favours exercise — Favours usual care

Analysis 2.2. Comparison 2 Exercise vs usual care, Outcome 2 Pain.

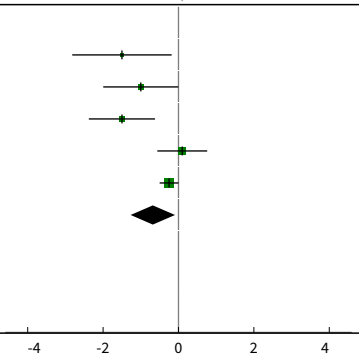
Study or subgroup	Exercise N	Exercise Mean(SD)	Usual care N	Usual care Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
2.2.1 End of intervention							
Rodriguez-Lozano 2013	381	-0.7 (2.2)	375	-0.4 (2.2)		82.95%	-0.18[-0.32,-0.03]
Sweeney 2002	75	-0.3 (1.5)	80	-0.2 (1.5)		17.05%	-0.07[-0.38,0.25]
Subtotal ***	**456**		**455**			**100%**	**-0.16[-0.29,-0.03]**
Heterogeneity: Tau^2=0; Chi^2=0.39, df=1(P=0.53); I^2=0%							
Test for overall effect: Z=2.38(P=0.02)							



Favours exercise — Favours usual care

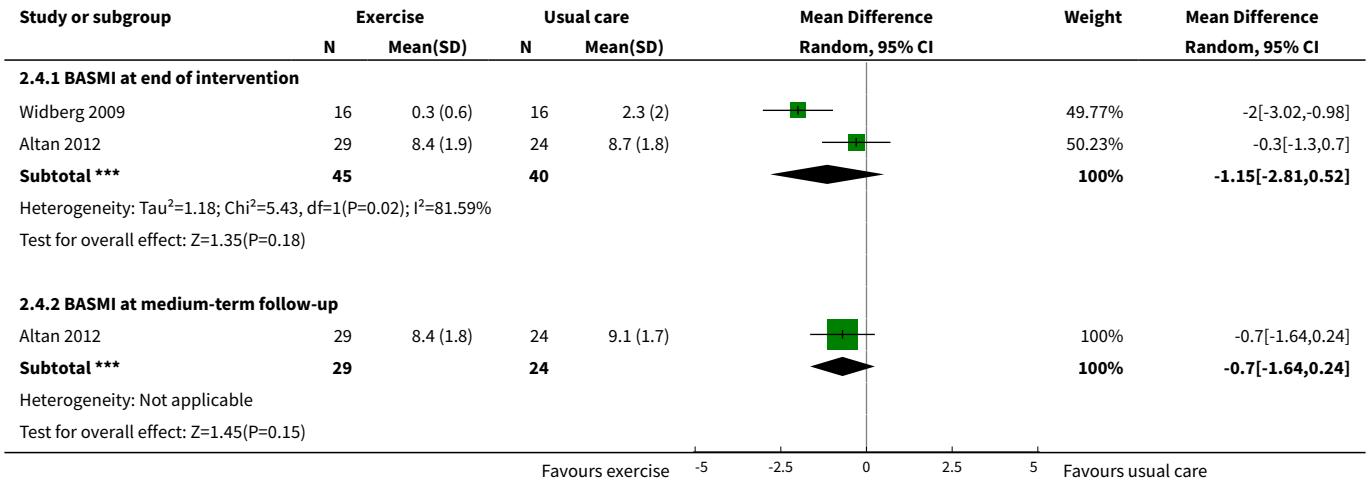
Analysis 2.3. Comparison 2 Exercise vs usual care, Outcome 3 Patient global assessment of disease activity.

Study or subgroup	Exercise N	Exercise Mean(SD)	Usual care N	Usual care Mean(SD)	Mean Difference Random, 95% CI	Weight	Mean Difference Random, 95% CI
2.3.1 BASDAI at end of intervention							
Widberg 2009	16	2.9 (1.8)	16	4.4 (2)		12.25%	-1.5[-2.82,-0.18]
Altan 2012	29	2.1 (2)	24	3.1 (1.7)		16.61%	-1[-2,-0]
Kjeken 2013	37	4.3 (1.9)	35	5.8 (1.9)		18.58%	-1.5[-2.38,-0.62]
Sweeney 2002	75	3.6 (2)	80	3.5 (2.2)		22.62%	0.1[-0.56,0.76]
Rodriguez-Lozano 2013	381	-0.6 (1.7)	375	-0.4 (1.8)		29.93%	-0.25[-0.5,-0]
Subtotal ***	**538**		**530**			**100%**	**-0.68[-1.27,-0.09]**
Heterogeneity: Tau^2=0.29; Chi^2=13.54, df=4(P=0.01); I^2=70.45%							
Test for overall effect: Z=2.26(P=0.02)							

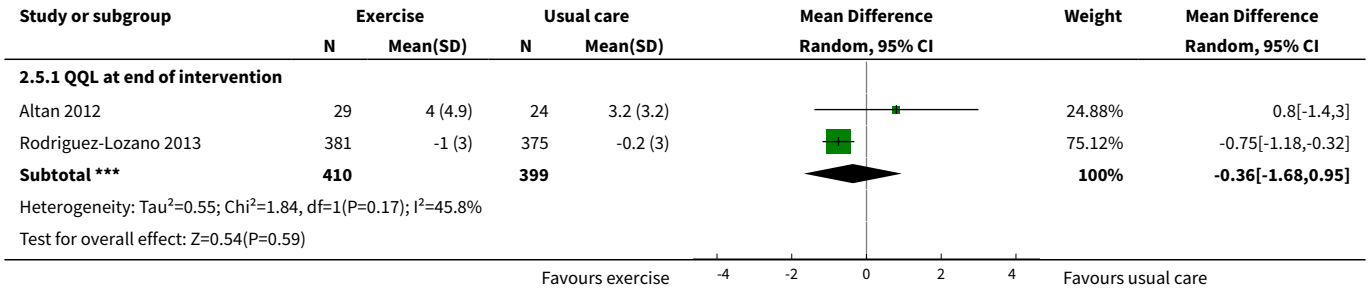


Favours exercise — Favours usual care

Analysis 2.4. Comparison 2 Exercise vs usual care, Outcome 4 Spinal mobility.



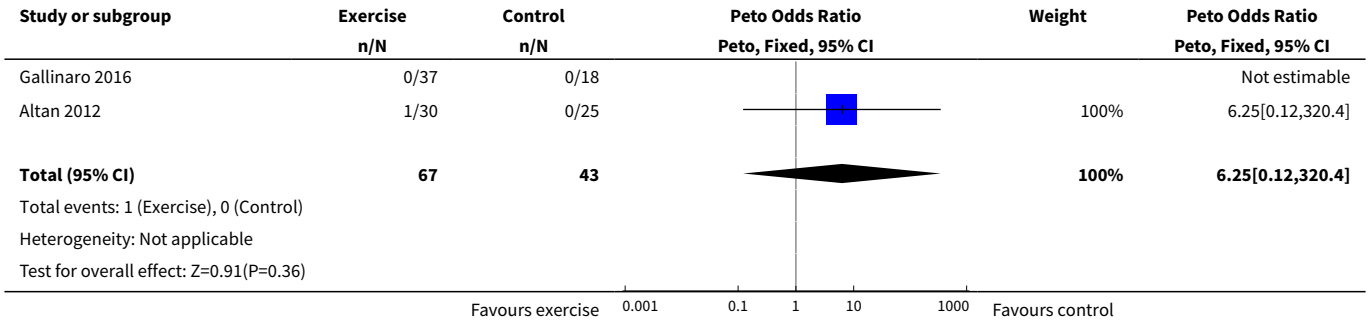
Analysis 2.5. Comparison 2 Exercise vs usual care, Outcome 5 Quality of life.



Comparison 3. Safety

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Adverse effects associated with the exercise intervention	2	110	Peto Odds Ratio (Peto, Fixed, 95% CI)	6.25 [0.12, 320.40]

Analysis 3.1. Comparison 3 Safety, Outcome 1 Adverse effects associated with the exercise intervention.



ADDITIONAL TABLES

Table 1. Authors contacted for missing or additional data

Authors	First contact	Second contact	Response
Altan 2012	25/05/2015	-	28/05/2015
Colina 2009	04/05/2015	13/05/2015	no e-mail response
Durmus 2009	15/04/2015	04/05/2015	no e-mail response
Dönmez 2014	15/04/2015	-	18/04/2015
Gunay 2012	15/04/2015	04/05/2015	no e-mail response
Ince 2006	02/06/2015	-	05/06/2015
Kjeken 2013	02/06/2015	-	03/06/2015
Kraag 1990	02/06/2015	-	03/06/2015
Lim 2005	no available contact	-	no e-mail response
Masiero 2011	20/04/2015	04/05/2015	04/05/2015
Masiero 2015	16/06/2016	-	20/06/2016
Mesquita 2014	15/04/2015	04/05/2015	no e-mail response
Rodriguez-Lozano 2013	02/06/2015	-	03/06/2015
Sveaas 2014	02/06/2015	-	04/06/2015
Sveeas 2018	24/01/2018		no e-mail response
Sweeney 2002	19/05/2015	02/06/2015	no e-mail response
Widberg 2009	02/06/2015	-	06/06/2015

Table 2. Summary of characteristics of included studies (N = 14)

Characteristics	N (%) or median (IQR)
Location	2 (14%)
Brazil	1 (7%)
Canada	1 (7%)
Italy	1 (7%)
Korea	2 (14%)
Norway	2 (14%)
Spain	1 (7%)
Sweden	3 (21%)
Turkey	1 (7%)
UK	
Study design	14 (100%)
RCT	
Number of study arms	11 (79%)
2	3 (21%)
3	
Type of comparator	5 (36%)
Usual care	9 (64%)
No treatment	
Total number participants per study	55 (35 to 73)
Trial size	3 (21)
> 100 subjects/arm	11 (79)
≤ 100 subjects/arm	
Number subjects per arm	26 (15 to 29)
Study duration (weeks)	14 (range 12 to 24)

N (%) is the number of studies that reported the characteristic of interest

Table 3. Summary of characteristics of participants in included studies (N = 14)

Characteristics	N (%) or median (IQR)
Age (years)	45 (39 to 47)
Gender	70 (56 to 77)

Table 3. Summary of characteristics of participants in included studies (N = 14) *(Continued)*

Male	33 (25 to 45)
Female	
Diagnostic criteria*	10 (71%)
Modified New York	2 (14%)
The Ankylosing Spondylitis Disease Activity Score	1 (7%)
European spondyloarthropathy	2 (14%)
not reported	
Severity disease*	5 (36%)
Bath Ankylosing Spondylitis Disease Activity Index ≥ 3.5	2 (14%)
Bath Ankylosing Spondylitis Disease Activity Index < 3.5	4 (29%)
Ankylosing Spondylitis stage1 or 2	3 (21%)
no information	
Disease duration (years)	9 (9 to 18)
coexisting medical treatments	21% (16% to 26%)
Analgesics (in 2 studies)	29% (14% to 38%)
Anti-Tumour Necrosis Factor (in 7 studies)	17% (11% to 19%)
Disease Modifying Anti-Rheumatic Drug (in 5 studies)	75% (32% to 76%)
Nonsteroidal anti-inflammatory drugs (in 9 studies)	22% (11% to 49%)
Sulfasalazine (in 4 studies)	17% (10% to 15%)
No treatment (in 2 studies)	NA
No information reported (in 4 studies)	

* N (%) is the number of studies that reported the characteristic of interest

Table 4. Summary of exercise programme characteristics in the included studies (N = 14)

Characteristics	N (%) or median (IQR)
Modalities	9 (64%)
Monomodal	5 (36%)
Multidisciplinary	
Exercise components	1 (7%)
Pain relief	7 (50%)
Breathing	2 (14%)
Cardio fitness	8 (57%)
Flexibility, stretching	1 (7%)

Table 4. Summary of exercise programme characteristics in the included studies (N = 14) *(Continued)*

Endurance	5 (36%)
Motion (active or passive)	4 (29%)
Proprioception, posture	2 (14%)
Relaxation	9 (64%)
Strength	1 (7%)
no information	
Provider	7 (50%)
Physiotherapist	3 (21%)
Other trainer	2 (14%)
Self delivery	2 (14%)
Unclear	
Supervision	8 (50%)
With supervision	3 (21%)
No supervision	3 (21%)
Unclear	
Dose	60 (50 to 60)
Session duration (minutes)	3 (2 to 3)
Frequency (session/week)	12 (8 to 16)
programme duration (weeks)	

N (%) is the number of studies that reported the characteristics of interest

Table 5. Major outcomes reported in the 14 included studies (part 1)

Study	Physical function (BASFI)	Patient global assessment (BASDAI)	Mobility (BASMI)	Mobility (chest expansion)	Mobility (occiput to wall distance)	Mobility (Schober test)	Mobility (Fingertip to floor)	Mobility (Cervical Rotation)
Altan 2012	Yes	Yes	Yes	Yes	-	-	-	-
Dönmez 2014	Yes	Yes	Yes	-	-	-	-	-
Garcia 2015	Yes†	Yes†	-	-	-	-	-	-
Gallinaro 2016	Yes ††	Yes ††	Yes ††	Yes ††		Yes ††	Yes ††	Yes
Ince 2006	-	-	-	Yes	Yes	Yes (modified)	Yes	-
Kjeken 2013	Yes	Yes	Yes*	-	-	-		-
Kraag 1990	-	-	-	-	Yes	Yes	Yes	-
Lim 2005	Yes	-	-	-	-	-	Yes	-
Masiero 2011	Yes	Yes	Yes	Yes	-	-	-	Yes
Rodriguez-Lozano 2013	Yes	Yes	-	-	-	-	-	-
Souza 2017	Yes	Yes	Yes	Yes	-	-	-	-
Sveaas 2014	Yes	Yes	Yes	-	-	-	-	-
Sweeney 2002	Yes	Yes	-	-	-	-	-	-
Widberg 2009	Yes	Yes	Yes	Yes	-	-	-	-

* Data are missing. cannot be included in the analysis
† median and 25th to 75th percentile reported
†† multiple exercise groups combined
BASFI: Bath Ankylosing Spondylitis Functionnal Index
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASMI: Bath Ankylosing Spondylitis Metrology Index

Table 6. Major outcomes reported in the 14 included studies (part 2)

Study	Pain (VAS)	Pain(SF-36)	Pain (BAS-DAI)	Pain (Nocturnal pain)	Pain (Self efficacy scale Pain)	Fatigue (Basdai)	Adverse Effects associated with exercise
Altan 2012	-	-	-	-	-		Yes
Dönmez 2014	Yes†	-	-	-	-		-
Garcia 2015	-	-	Yes	-	-	Yes	-
Gallinaro 2016	Yes ††	-	-	-	-	-	Yes ††
Ince 2006	-	-	-	-	-	-	-
Kjeken 2013	-	-	-	-	-	-	-
Kraag 1990	Yes	-	-	-	-	-	-
Lim 2005	Yes	-	-	-	-		-
Masiero 2011	Yes**	-	-	-	-	Yes	-
Rodriguez-Lozano 2013	Yes	-	-	Yes	-	-	-
Souza 2017	-	Yes	-	-	-	-	-
Sveaas 2014	-	-	-	-	-	-	-
Sweeney 2002	-	-	-	-	Yes	-	-
Widberg 2009	-	-	-	-	-	-	-

** mean score calculated from lumbar and cervical pain
† median and 25th to 75th percentile reported
†† multiple exercise groups combined
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
VAS: visual analogue scale
SF-36: 36-Item Short-Form Health Survey

Table 7. Minor outcomes reported in the 14 included studies

Study	Quality of life (ASQoL)	Quality of life (SF-36)	Quality of life (SF-12) physical component	CRP level (mg/dL)	ESR (mm/h)	MASES
Altan 2012	Yes	-	-	-	-	-
Dönmez 2014	-	Yes*	-	-	-	-
Garcia 2015	-	-	Yes†	-	-	-
Gallinaro 2016	-	-	Yes ††	-	-	Yes ††
Ince 2006	-	-	-	-	-	-
Kjeken 2013	-	Yes*	-	-	-	-
Kraag 1990	-	-	-	-	-	-
Lim 2005	-	-	-	-	-	-
Masiero 2011	-	-	-	not reported	not reported	-
Rodriguez-Lozano 2013	Yes	-	-	-	-	-
Souza 2017	-	Yes*	-	Yes	Yes	-
Sveaas 2014	-	-	-	Yes	Yes	-
Sweeney 2002	-	-	-	-	-	-
Widberg 2009	-	-	-	-	-	-

* global score was not reported; could not be included in the analysis
† median and 25th to 75th percentile reported
†† multiple exercise groups combined
ASQoL: the Ankylosing Spondylitis Quality of Life
SF-36: 36-Item Short-Form Health Survey
SF-12: 12-Item Short Form Health Survey
CRP: C-reactive protein
ESR: erythrocyte sedimentation rate
MASES: Maastricht Ankylosing SpondylitisEnthesitis Score

APPENDICES

Appendix 1. MEDLINE Ovid search strategy

1 Spondylitis, Ankylosing/

2 Spondylarthritis/)

3 (axial adj2 spondylarthritis).tw.

4 (axial adj2 spa).tw.

5 (ankylos$ or spondyl$).tw.

6 (bekhterev$ or bechterew$).tw.

7 (Marie adj struempell$).tw.

8 (AS or axSPA).tw.

9 Sacroiliitis/

10 ((axial or spin$ or peripheral or vertebral or enthesitis) adj3 (joint$ or spondyloarthritis or arthritis or ankylosing)).tw.

11 or/1-10

12 exp Exercise/

13 Physical Exertion/

14 Physical Fitness/

15 exp Physical Endurance/

16 exp Sports/

17 Pliability/

18 exertion$.tw.

19 exercis$.tw.

20 sport$.tw.

21 ((physical or motion) adj5 (fitness or therapy or therapies)).tw.

22 (physical$ adj2 endur$).tw.

23 manipulat$.tw.

24 (skate$ or skating).tw.

25 jog$.tw.

26 swim$.tw.

27 bicycl$.tw.

28 (cycle$ or cycling).tw.

29 walk$.tw.

30 (row or rows or rowing).tw.

31 weight train$.tw.

32 muscle strength$.tw.

33 exp Yoga/

34 yoga.tw.

35 exp Tai Ji/

36 tai chi.tw.

37 Ai Chi.tw.

38 exp Vibration/

39 vibration.tw.

40 pilates.tw.

41 Motor Activity/

42 exp Exercise Therapy/

43 exp Proprioception/

44 exp Physical Therapy Modalities/

45 exp Rehabilitation/

46 or/12-45

47 11 and 46

48 randomized controlled trial.pt.

49 controlled clinical trial.pt.

50 randomized.ab.

51 placebo.ab.

52 clinical trials as topic.sh.

53 randomly.ab.

54 trial.ti.

55 or/48-54

56 exp animals/ not humans.sh.

57 55 not 56

58 47 and 57

Appendix 2. CENTRAL search strategy

1. MeSH descriptor: [Exercise] explode all trees

2. Exert:ti,ab

3. MeSH descriptor: [Physical Fitness] explode all trees

4. MeSH descriptor: [Exercise Test] explode all trees

5. MeSH descriptor: [Exercise Tolerance] explode all trees

6. MeSH descriptor: [Sports] explode all trees

7. MeSH descriptor: [Pliability] explode all trees

8. MeSH descriptor: [Physical Endurance] explode all trees

9. exertion*:ti,ab

10. exercis*:ti,ab

11. sport*:ti,ab

12. ((physical or motion) near/5 (fitness or therap*)):ti,ab

13. (physical* near/2 endur*):ti,ab

14. ((strength* or isometric* or isotonic* or isokinetic* or aerobic* or endurance or weight*) near/5 (exercis* or train*)):ti,ab

15. MeSH descriptor: [Physical Therapy Modalities] explode all trees

16. (physical next therap*):ti,ab

17. physiotherap*:ti,ab

18. manipulat*:ti,ab

19. kinesiotherap*:ti,ab

20. MeSH descriptor: [Rehabilitation] explode all trees

21. rehab*:ti,ab

22. (skate* or skating):ti,ab

23. run*:ti,ab

24. jog*:ti,ab

25. treadmill*:ti,ab

26. swim*:ti,ab

27. bicycl*:ti,ab

28. (cycle* or cycling):ti,ab

29. walk*:ti,ab

30. (row or rows or rowing):ti,ab

31. muscle next strength:ti,ab

32. (#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31)

33. MeSH descriptor: [Spondylitis, Ankylosing] explode all trees

34. MeSH descriptor: [Spondylarthropathies] explode all trees

35. ankylosing or spondyl:ti,ab

36. bekhterev or bechterew:ti,ab

37. ((axial or spin* or peripheral or vertebral or enthesitis) near/3 (joint* or spondyloarthritis or arthritis or ankylosing)) .tw.

38. #33 or #34 or #35 or #36 or #37

39. #32 and #38

Appendix 3. Embase search strategy

1 ankylosing spondylitis/

2 (ankylos$ or spondyl$).tw.

3 (bekhterev$ or bechterew$).tw.

4 (Marie adj struempell$).tw.

5 sacroiliitis/

6 ((axial or spin$ or peripheral or vertebral or enthesitis) adj3 (joint$ or spondyloarthritis or arthritis or ankylosing)).tw.

7 (AS or axSPA).tw.

8 or/1-7

9 exp EXERCISE/

10 fitness/

11 exercise test/

12 exercise tolerance/

13 exp Sport/

14 pliability/

15 exp "physical activity, capacity and performance"/

16 exertion$.tw.

17 exercis$.tw.

18 sport$.tw.

19 ((physical or motion) adj5 (fitness or therap$)).tw.

20 (physical$ adj2 endur$).tw.

21 ((strength$ or isometric$ or isotonic$ or isokinetic$ or aerobic$ or endurance or weight$) adj5 (exercis$ or train$)).tw.

22 exp physiotherapy/

23 physiotherap$.tw.

24 manipulat$.tw.

25 kinesiotherap$.tw.

26 exp REHABILITATION/

27 rehab$.tw.

28 (skate$ or skating).tw.

29 run$.tw.

30 jog$.tw.

31 treadmill$.tw.

32 swim$.tw.

33 bicycl$.tw.

34 (cycle$ or cycling).tw.

35 walk$.tw.

36 (row or rows or rowing).tw.

37 muscle strength$.tw.

38 or/9-37

39 random$.ti,ab.

40 factorial$.ti,ab.

41 (crossover$ or cross over$ or cross-over$).ti,ab.

42 placebo$.ti,ab.

43 (doub$ adj blind$).ti,ab.

44 (singl$ adj blind$).ti,ab.

45 assign$.ti,ab.

46 allocat$.ti,ab.

47 volunteer$.ti,ab.

48 crossover procedure.sh.

49 double blind procedure.sh.

50 randomized controlled trial.sh.

51 single blind procedure.sh.

52 or/39-51

53 exp animal/ or nonhuman/ or exp animal experiment/

54 exp human/

55 53 and 54

56 53 not 55

57 8 and 38 and 56

Appendix 4. CINAHL search strategy

S40	S7 AND S39
S39	S38 or S37 or S36 or S35 or S34 or S33 or S32 or S31 or S30 or S29 or S28 or S27 or S26 or S25 or S24 or S23 or S22 or S21 or S20 or S19 or S18 or S17 or S16 or S15 or S14 or S13 or S12 or S11 or S10 or S10 or S9 or S8
S38	(ti "muscle strength*") or (ab "muscle strength*")
S37	(ti row or rows or rowing) or (ab row or rows or rowing)
S36	(ti walk*) or (ab walk*)
S35	(ti cycle* or cycling) or (ab cycle* or cycling)
S34	(ti bicycl*) or (ab bicycl*)
S33	(ti swim*) or (ab swim*)
S32	(ti swim*) or (ab swim*)
S31	(ti treadmill*) or (ab treadmill*)
S30	(ti jog*) or (ab jog*)
S29	(ti run*) or (ab run*)
S28	(ti skate* or skating) or (ab skate* or skating)
S27	(ti rehab*) or (ab rehab*)
S26	(MH "Rehabilitation+")

(Continued)

S25	(ti kinesiotherap*) or (ab kinesiotherap*)
S24	(ti manipulat*) or (ab manipulat*)
S23	(ti physiotherap*) or (ab physiotherap*)
S22	(MH "Physical Therapy+")
S21	TI (strength* or isometric* or isotonic* or isokinetic*or aerobic* or endurance or weight*) or AB (strength* or isometric* or isotonic* or isokinetic*or aerobic* or endurance or weight*)
S20	TI physical* n2 endur* or AB physical* n2 endur*
S19	TI physical N5 fitness or TI physical N5 therap* or AB physical N5 fitness or AB physical N5 therap* or TI motion n5 therap* or AB motion n5 therap*
S18	(ti sport*) or (ab sport*)
S17	(ti exercis*) or (ab exercis*)
S16	(ti exertion*) or (ab exertion*)
S15	(MH "Physical Endurance+")
S14	(MH "Pliability")
S13	(MH "Sports+")
S12	(MH "Exercise Tolerance+")
S11	(MH "Exercise Test+")
S10	(MH "Physical Fitness")
S9	(MH "Exertion+")
S8	(MH "Exercise+")
S7	S1 OR S2 OR S3 OR S4 OR S5 OR S6
S6	(MH "Spondylarthritis")
S5	"Sacroiliitis"
S4	"Axial Spondyloarthritis"
S3	"AS or axSPA"
S2	"(bekhterev or bechterew)"
S1	(MH "Spondylitis, Ankylosing")

Appendix 5. PEDro search strategy

Topic: spondylitis or spondyloarthritis

Intervention: Fitness training, Strength training, Stretching, mobilisation, manipulation, massage

Appendix 6. Clinicaltrials.gov search strategy

all years

Ankylosing spondylitis in Condition

Exercise in Intervention

Appendix 7. WHO ICTRP search strategy

Ankylosing spondylitis in Condition
Exercise in Intervention

WHAT'S NEW

Date	Event	Description
28 January 2020	Amended	Corrected SoF table 1

CONTRIBUTIONS OF AUTHORS

JPR, MMLC, CP, SP, TD, and IB contributed to the development of the protocol. All authors were involved in the conception and interpretation of the review.

DECLARATIONS OF INTEREST

JPR: none known

MMLC: none known

CP: received consultancy remuneration from Merz, Novartis, Ipsen, and support for travel from Merz; received remuneration from BMS for participating in Delphi study to develop a questionnaire on fear, beliefs, and expectations regarding pain induced by physiotherapy, and from Pfizer for the analysis of a national survey in 2010-2011. Pfizer, BMS, Roche contributed to expenses to conferences (EULAR 2011 and 2012, SFR 2012).

AR: none known

FR: declares competing interest outside the submitted work: board membership of Pfizer, Sanofi Aventis, Pierre Fabre, Expansciences and Thuasne. Consultancy for Genevrier and Bayer. Payment for lectures from Thuasne, Grünenthal and IPSEN.

TD: association Robert Debré provided a grant for master studies; he works as a physiotherapist.

IB: is a co-convener of the Bias Methods Group (BMG) of the Cochrane Collaboration and the French satellite co-ordinator of the Cochrane Musculoskeletal Review Group (CMSG)

SOURCES OF SUPPORT

Internal sources

- French School of Public Health (EHESP) Paris Sorbonne Cité, France.

  in-kind support
- Centre de Recherche Inserm Epidémiologie et Statistique Paris Sorbonne Cité (CRESS), France.

  in-kind support
- Hôpital Hôtel-Dieu, APHP, France.

  in-kind support

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

1. We did not perform contour-enhanced funnel plots to assess the presence of small-study effects because the required statistical conditions were not met.
2. We did not perform subgroup analysis to explore a relationship between the modalities of exercises, participant characteristics, or pharmacological treatments because of insufficient data. Meta-regression (i.e. dose–effect relationship) was also not possible because of the small number of included studies for each reported outcome.
3. The protocol stated that we would attempt to perform a sensitivity analysis to explore how the results of meta-analysis might be affected by including only studies at low risk of bias. However, because all the identified studies were at high risk of bias, we did not perform the analysis.
4. We used Peto odds ratios (Peto ORs) for calculating AEs because too few events were reported in each group (Higgins 2011b).
5. We planned in the protocol to report outcome assessments from the included studies in three time frames: at completion (end of intervention), medium-term follow-up (6 to 12 months), and long-term follow-up (> 12 months). We revised the time frames in the review to: at completion (end of intervention), medium-term (less than six months after completion of exercise), and long-term (6 months and more after completion of exercise) follow-up.
6. In the protocol, we named the outcome to measure the patient-assessment of disease activity, 'Patient global assessment of health status'. We changed the name to 'Patient global assessment of disease activity' in the review
7. François Rannou and Alexandra Roren were added as authors.